



Essays on Trade and Imperfectly Competitive Markets

Citation

Fabinger, Michal. 2012. Essays on Trade and Imperfectly Competitive Markets. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:9548614>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

©2012 — *Michal Fabinger*

All rights reserved.

Primary Dissertation Advisor:
Gita Gopinath

Michal Fabinger

Essays on Trade and Imperfectly Competitive Markets

ABSTRACT

This dissertation is a collection of three essays on markets with imperfect competition, with implications for international economics.

The first essay presents an analytic solution framework applicable to a wide variety of general equilibrium international trade models, including those of Krugman (1980), Eaton and Kortum (2002), Anderson and van Wincoop (2003), and Melitz (2003), in multi-location cases. For asymptotically power-law trade costs and in the large-space limit, it is shown that there are parameter thresholds where the qualitative behavior of the model economy changes. In the case of the Krugman (1980) model, the relevant parameter is closely related to the elasticity of substitution between different varieties of goods. The geographic reach of economic shocks changes fundamentally when the elasticity crosses a critical threshold: below this point shocks are felt even at long distances, while above it they remain local. The value of the threshold depends on the approximate dimensionality of the spatial configuration. This work bridges the gap between empirical work on international and intranational trade, which frequently uses data sets involving large numbers of locations, and the theoretical literature, which has analytically examined solutions to the relevant models with realistic trade costs only for the case of very few locations.

The second essay, coauthored with Glen Weyl, extends the incidence-based frame-

work for the analysis of perfectly competitive markets to imperfect competition. We show how, just as under perfect competition, a wide range of comparative statics and policy analyses turn on simple properties of incidence, particularly the rate at which unit taxes are passed through to consumer prices. We derive local and global incidence properties, the division of surplus among deadweight loss, consumer surplus and profits and show how these are linked to one another under a range of imperfectly competitive environments. We then show how incidence functions as a simplifying analytic and pedagogic device, an empirical sufficient statistic and a key structural parameter in both classic and recently popular topics in industrial economics including platforms, concession auctions, mergers, entry, price discrimination, product design, supply chains and advertising.

The third essay, coauthored with Gita Gopinath and Oleg Itskhoki, studies pricing of durable goods by producers with market power. The durable nature of these products makes their pricing differ from that of nondurables, since consumer demand depends not only on prices today but also on their expectation of future prices. When firms cannot commit to future prices, pass-through of cost shocks into prices is incomplete and the adjustment is gradual. This is the case even when prices are fully flexible and in environments where non-durable pricing would generate complete pass-through. Prices are also sensitive to demand shocks and mark-ups are pro-cyclical, in contrast to the case of cost shocks when mark-ups are countercyclical. We present these results for the case of a monopolist, for oligopolistic competition and for monopolistic competition.

TABLE OF CONTENTS

| | |
|--|-----|
| <i>Abstract</i> | iii |
| <i>Acknowledgments</i> | ix |
| <i>Introduction</i> | 1 |
| 1. <i>Trade and Interdependence in a Spatially Complex World</i> | 4 |
| 1.1 Introduction | 4 |
| 1.2 Challenges of multi-location models | 9 |
| 1.2.1 Working with only a few locations | 10 |
| 1.2.2 Neglecting changes in general equilibrium effects | 13 |
| 1.3 The Krugman model | 13 |
| 1.3.1 Production and transportation | 13 |
| 1.3.2 Consumption | 14 |
| 1.3.3 Closing the model | 15 |
| 1.3.4 Comparative statics - part 1 | 16 |
| 1.3.5 The GDP propagator | 18 |
| 1.3.6 Comparative statics - part 2 | 20 |
| 1.4 The Krugman model in continuous space | 21 |
| 1.4.1 Change in the solution in response to a small change in trade costs | 23 |
| 1.4.2 Welfare | 26 |
| 1.4.3 Asymptotically power-law transportation costs | 27 |
| 1.5 The Krugman model on the circle | 29 |
| 1.5.1 Basic setup | 29 |
| 1.5.2 Expansion in terms of convolution powers of $G_c(\theta)$ | 30 |
| 1.5.3 General solution for y_1 and $y_1^{(P)}$ | 36 |
| 1.5.4 Fourier coefficients of $G_c(\theta)$ for specific functional forms of trade costs | 38 |
| 1.6 The impact of border costs | 39 |
| 1.6.1 General solution for GDP in the presence of border costs | 39 |
| 1.6.2 Bounds on $y_1^{(P)}(0) / y_1^{(P)}(\frac{\pi}{2})$ | 42 |

| | | |
|-------|---|-----|
| 1.7 | The impact of changes in productivity | 42 |
| 1.8 | The Krugman model on the sphere | 46 |
| 1.8.1 | The role of dimensionality | 46 |
| 1.8.2 | Basic setup | 46 |
| 1.8.3 | General solution for y_1 and $y_1^{(P)}$ | 48 |
| 1.8.4 | Solutions for specific functional forms of trade costs | 50 |
| 1.8.5 | The impact of changes in border costs | 51 |
| 1.8.6 | The impact of changes in productivity | 51 |
| 1.9 | Higher-order terms | 52 |
| 1.10 | Conclusion | 55 |
| 2. | <i>Pass-through as an Economic Tool</i> | 56 |
| 2.1 | Introduction | 56 |
| 2.2 | Pass-through | 60 |
| 2.2.1 | Examples: Cournot and differentiated Bertrand competition | 61 |
| 2.2.2 | Pass-through formula | 62 |
| 2.2.3 | Discussion | 64 |
| 2.3 | Local and global incidence | 67 |
| 2.3.1 | Global incidence | 68 |
| 2.3.2 | Excess burden | 70 |
| 2.4 | Graphical illustration | 72 |
| 2.5 | Platforms | 74 |
| 2.5.1 | The Rochet and Tirole (2003) model | 74 |
| 2.5.2 | The Becker (1991) model | 79 |
| 2.6 | Procuring new markets | 81 |
| 2.6.1 | Concession auctions | 81 |
| 2.6.2 | Innovation incentives | 83 |
| 2.6.3 | Project choice/design | 85 |
| 2.7 | Other Applications | 87 |
| 2.7.1 | Mergers | 87 |
| 2.7.2 | Entry | 89 |
| 2.7.3 | Price discrimination | 90 |
| 2.7.4 | Product design | 93 |
| 2.7.5 | Supply chains | 96 |
| 2.7.6 | Aftermarkets and advertising | 97 |
| 2.7.7 | Taxing foreign trade | 99 |
| 2.7.8 | Demand forms | 100 |
| 2.7.9 | An empirical example | 101 |
| 2.8 | Conclusion | 102 |

| | | |
|-------|---|-----|
| 3. | <i>Price Dynamics for Durable Goods</i> | 105 |
| 3.1 | Introduction | 105 |
| 3.2 | Durable good demand | 109 |
| 3.3 | Durable good monopoly | 112 |
| 3.3.1 | The commitment case as a benchmark | 112 |
| 3.3.2 | Discretionary pricing of durable goods | 115 |
| 3.3.3 | Extension: Calvo sticky price setting | 129 |
| 3.4 | Durable good oligopoly | 131 |
| 3.5 | Durable good monopolistic competition | 136 |
| 3.6 | Conclusion | 142 |
| | <i>Appendix</i> | 143 |
| A. | <i>Appendix to Chapter 1</i> | 144 |
| A.1 | Neglecting changes in general equilibrium effects | 144 |
| A.2 | Remarks on methodology: reverse engineering equilibria from comparative statics | 146 |
| A.2.1 | The case of a single endogenous variable | 146 |
| A.2.2 | The case of two endogenous variables and its generalization | 149 |
| A.3 | Derivation of equation (1.18) | 152 |
| A.4 | Local-price-index-adjusted GDP | 153 |
| A.5 | Properties of $G_c^{*2}(\theta)$ on the circle for large R | 154 |
| A.5.1 | The peaks of $G_c^{*2}(\theta)$ | 156 |
| A.5.2 | $\delta < \frac{1}{2}$, tails of $G_c^{*2}(\theta)$ | 157 |
| A.5.3 | $\delta > \frac{1}{2}$, tails of $G_c^{*2}(\theta)$ | 158 |
| A.6 | Properties of $G_c^{*2}(\theta)$ on the sphere for large R | 159 |
| A.6.1 | The peaks of $G_c^{*2}(\theta)$ | 161 |
| A.6.2 | $\delta < 1$, tails of $G_c^{*2}(\theta)$ | 162 |
| A.6.3 | $\delta > 1$, tails of $G_c^{*2}(\theta)$ | 163 |
| A.7 | Fourier series expansions | 165 |
| A.7.1 | Fourier series expansions of country indicator functions | 165 |
| A.7.2 | Fourier expansion of $\tilde{h}(\theta)$ | 166 |
| A.7.3 | Fourier expansion of $\tilde{g}_c(\theta)$ | 169 |
| A.8 | Derivation of the expression for T_n on circle | 169 |
| A.9 | Spherical harmonic expansions | 171 |
| A.9.1 | Spherical harmonic expansions of country indicator functions | 171 |
| A.9.2 | Spherical harmonic expansion of $Y_{l'}^0(\theta, \varphi) C_B(\theta, \varphi)$ and $Y_{l'}^0(\theta, \varphi) C_A(\theta, \varphi)$ | 172 |
| A.9.3 | Spherical harmonic expansions used to analyze the impact of border costs | 174 |

| | | |
|--------|--|-----|
| A.9.4 | Spherical harmonic expansion of $\tilde{g}_c(x)$ | 176 |
| A.10 | Derivation of the expression for T_l for the sphere | 177 |
| A.11 | Relation to fields in anti de Sitter space | 180 |
| A.12 | The impact of border costs in the large-space limit on the circle | 182 |
| A.12.1 | Nonnegativity of $(-1)^{\frac{n}{2}} \tilde{g}_{c,n}$ | 182 |
| A.12.2 | A lower bound on $-y_1^{(P)}\left(\frac{\pi}{2}\right)$ | 183 |
| A.12.3 | An upper bound on $-y_1^{(P)}\left(\frac{\pi}{2}\right)$ | 183 |
| A.12.4 | A lower bound on $-\lim_{R \rightarrow \infty} y_1^{(P)}(0)$ for $\delta < \frac{1}{2}$ | 184 |
| A.12.5 | An alternative derivation of the bounds (A.28) and (A.29) in the large-space limit for $\delta > \frac{1}{2}$ | 190 |
| A.12.6 | Evaluating $\lim_{R \rightarrow \infty} y^{(P)}(0)$ for $\delta > \frac{1}{2}$ | 197 |
| A.12.7 | Conclusion | 198 |
| A.13 | Mathematical notation | 200 |
| B. | <i>Appendix to Chapter 2</i> | 201 |
| C. | <i>Appendix Chapter 3</i> | 204 |
| | <i>Bibliography</i> | 206 |

ACKNOWLEDGMENTS

I owe enormous debts to my dissertation committee of Pol Antràs, Gita Gopinath, Marc Melitz, and Glen Weyl. My primary advisor Gita Gopinath provided extremely insightful guidance for my dissertation, and she collaborated with Oleg Itskhoki and myself on the work reported in Chapter 3. In addition, she selflessly spent many hours discussing with me various topics in international economics not directly related to our research in order to help me broaden my knowledge of the research field. Insights from Pol Antràs and Marc Melitz were crucial in shaping the project reported in Chapter 1. They helped me understand the wide applicability of the results, and they gave me important advice on writing the corresponding research paper. Glen Weyl collaborated with me on the material presented in Chapter 2 and on other projects and gave me invaluable advice on all parts of the dissertation. In addition, he frequently discussed with me interesting topics in many different fields of economics and in related disciplines.

The research presented in this dissertation also benefitted from extremely helpful discussions with Mark Aguiar, Manuel Amador, James Anderson, Jeremy Bulow, Lorenzo Caliendo, Luciano de Castro, Arnaud Costinot, Dave Donaldson, Claudia Goldin, Elhanan Helpman, Oleg Itskhoki, Kevin Murphy, Nathan Nunn, Ariel Pakes, Rob Porter, Michael Salinger, and Robert Townsend. Ahmed Jaber, Rosen Kralev, Stephanie Lo, Yali Miao, Toan Phan, and William Weingarten contributed as research assistants to the project of Chapter 2.

I was very fortunate to meet Roland Fryer and Ben Olken, who inspired me to learn economics. Support from Edward Glaeser, Jerry Green, Michael Kremer, Send-

hil Mullainathan, and Geert Ridder was crucial for entering the economics graduate program at Harvard University. I am thankful to Wally Gilbert for his encouragement to pursue economics, and to Diana Morse who helped me in many practical respects.

I am grateful for all the help that I received from my family and my friends. This dissertation would not have been written without the support of my wife Mika Iga, who, despite her long working hours, had the courage to let me pursue studies of economics at a time when our sons Alan and Luca required much attention from their parents.

INTRODUCTION

In this dissertation, I bring together three essays on markets with imperfect competition, with implications for international economics. The first essay describes novel solutions to most widely used models of international trade and highlights the importance of geography for transmission of economic shocks. The second essay provides a new perspective on imperfectly competitive markets, with the pass-through rate playing a central role. This approach streamlines welfare analysis in a wide variety of economic models and makes it easier to assess robustness of the models' predictions. The third essay studies pricing of durable goods by producers with market power who are unable to commit to high prices in the future. The essay analyzes the implications for business-cycle fluctuations and for cost pass-through (e.g., exchange rate pass-through).

Chapter 1 presents an analytic solution framework applicable to a wide variety of general equilibrium international trade models, including those of Krugman (1980), Eaton and Kortum (2002), Anderson and van Wincoop (2003), and Melitz (2003), in multi-location cases. For asymptotically power-law trade costs and in the large-space limit, it is shown that there are parameter thresholds where the qualitative behavior of the model economy changes. In the case of the Krugman (1980) model, the relevant parameter is closely related to the elasticity of substitution between different varieties of goods. The geographic reach of economic shocks changes fundamentally when the elasticity crosses a critical threshold: below this point shocks are felt even at long

distances, while above it they remain local. The value of the threshold depends on the approximate dimensionality of the spatial configuration. This work bridges the gap between empirical work on international and intranational trade, which frequently uses data sets involving large numbers of locations, and the theoretical literature, which has analytically examined solutions to the relevant models with realistic trade costs only for the case of very few locations.

Chapter 2, coauthored with Glen Weyl, extends the incidence-based framework for the analysis of perfectly competitive markets to imperfect competition. We show how, just as under perfect competition, a wide range of comparative statics and policy analyses turn on simple properties of incidence, particularly the rate at which unit taxes are passed through to consumer prices. We derive local and global incidence properties, the division of surplus among deadweight loss, consumer surplus and profits and show how these are linked to one another under a range of imperfectly competitive environments. We then show how incidence functions as a simplifying analytic and pedagogic device, an empirical sufficient statistic and a key structural parameter in both classic and recently popular topics in industrial economics including platforms, concession auctions, mergers, entry, price discrimination, product design, supply chains and advertising.

Chapter 3, coauthored with Gita Gopinath and Oleg Itskhoki, studies pricing of durable goods by producers with market power. Durable goods, of course, are known to represent a large fraction of international trade flows. The durable nature of these products makes their pricing differ from that of nondurables, since consumer demand depends not only on prices today but also on their expectation of future prices. When firms cannot commit to future prices, pass-through of cost shocks into prices is incomplete and the adjustment is gradual. This is the case even when prices are fully

flexible and in environments where non-durable pricing would generate complete pass-through. Prices are also sensitive to demand shocks and mark-ups are procyclical, in contrast to the case of cost shocks when mark-ups are countercyclical. We present these results for the case of a monopolist, for oligopolistic competition, and for monopolistic competition. In the case of monopolistic competition studied here, price dynamics is governed both by the dynamics of the stock of individual varieties relative to the industry average and by the movements in the industry average, with more weight placed on the latter.

1. TRADE AND INTERDEPENDENCE IN A SPATIALLY COMPLEX WORLD

1.1 *Introduction*

Suppose that the cost of moving goods across the border between two large countries changes. How far from the border is the economic impact going to be felt? Do such changes mostly affect regions close to the border, or do they significantly affect even very distant locations? What if productivity increases or decreases in one of these countries, due to an economic boom or due to a crisis? How is the productivity change going to influence the level of welfare at various places in the other country?

To address these questions, it is natural to employ standard models of international trade, such as Krugman (1980). The solutions to these models have been theoretically analyzed for some cases. If there are just two or three locations where economic activity takes place, the analysis is very straightforward.¹ To gain insight into situations with many locations, the theoretical literature has used certain analytically convenient specifications of trade costs. Apart from zero trade costs, the most popular assumption corresponds to ‘symmetric trade costs’, in which case the cost of trade between any pair of distinct locations is the same.² For example, the multilat-

¹ Matsuyama (1999) solves interesting cases with as many as eight locations in the context of the model introduced in Section 10.4 of Helpman and Krugman (1987), which adds a costlessly tradable homogeneous good to Krugman (1980).

² In the context of economic geography models (see Fujita, Krugman, and Venables (1999)), trade costs

eral trade policy analysis in Baldwin, Forslid, Martin, Ottaviano, and Robert-Nicoud (2005) builds on this assumption.

For the present purposes, however, it is necessary to work in a multi-location setting with more realistic trade costs. Clearly, the transportation costs should grow with distance. At the same time, they should reflect economies associated with shipping goods over long distances: the per-unit-distance transportation cost should be a decreasing function of distance.³

The empirical literature has been working with trade models at this level of realism for a long time. In recent years, the multi-location aspect has become prominent in empirical work. Due to falling costs of information technology, highly spatially disaggregated data sets are becoming available for empirical analysis. For example, Hillberry and Hummels (2008) study manufacturers' shipments within the United States with 5-digit zip code precision. Compared to previous studies this is a remarkable improvement in spatial resolution.

The aim of the present work is to bridge the gap between the context in which international trade models are used for empirical purposes and the context in which they are studied theoretically. The article introduces a mathematical framework⁴ that allows one to solve and analyze such trade models in basic cases involving many locations. The model discussed extensively is that of Krugman (1980), but this choice is made primarily for expositional purposes. The models of Anderson and van Win-

exponential in distance proved to be a convenient choice. In that specification, the per-unit-distance trade cost is an increasing function of distance.

³ See Anderson and van Wincoop (2004) and Hummels (2001) for empirical evidence on trade costs.

⁴ The framework makes extensive use of standard tools of functional analysis. In the concrete examples considered, these are Fourier series expansion and spherical harmonic expansion.

coop (2003)/Armington (1969)⁵ or Melitz (2003)⁶ have a very similar mathematical structure,⁷ and only minor modifications are needed to write down their solutions once the solutions to the Krugman model are known. The same is true⁸ for the Ricardian model of Eaton and Kortum (2002). This method may also be applied to many other types of trade models, such as those in Baldwin, Forslid, Martin, Ottaviano, and Robert-Nicoud (2005), where some factors of production are frequently assumed to be mobile.

What are the practical lessons coming from the analysis? Take the Krugman model as a representative example. Let the transportation costs be of the ‘iceberg’ type and asymptotically power-law⁹ in distance, as commonly assumed in the empirical literature. Suppose also that the spatial geometry is very large and homogeneously populated. In this case, it turns out, the way general equilibrium effects spread through the economy depends very strongly on the elasticity of substitution between different varieties of goods. When the elasticity is above a certain threshold, disturbances spread through the economy by short-distance interactions. With the elasticity below the threshold, interactions between economic agents separated by long distances play a crucial role. This fact has important consequences for various quantities of interest.

Consider the case of two large neighboring countries mentioned earlier, and suppose that the cost associated with moving goods across the border increases slightly.

⁵ The model of Anderson and van Wincoop (2003) is an extension of Armington (1969) and Anderson (1979).

⁶ For a derivation of the corresponding gravity equation, see Chaney (2008).

⁷ See Arkolakis, Costinot, and Rodríguez-Clare (forthcoming) for a detailed analysis of the similarities between the models.

⁸ Also the portfolio choice model of Okawa and van Wincoop (2010) has the same property.

⁹ For a clarification of the term ‘asymptotically power-law,’ see Subsection 1.4.3.

If the elasticity is above the threshold, only locations close to the border will be affected. On the other side of the threshold, the change in the border cost significantly affects all locations. In the case of a productivity change in one of the countries, the situation is similar. With the elasticity above the threshold, the effects on the other country will be restricted to a small region close to the border. For the elasticity below the threshold, the consequences of the productivity change will be felt throughout the other country.

At the empirical level, these observations imply that when fitting a similar trade model to the data, the usual practice of assuming that all differentiated goods have the same elasticity of substitution can lead to unexpectedly strong biases. The properties of the model are highly non-linear in the elasticity. Under such circumstances, replacing heterogeneous goods with a single type of good having the average elasticity is misleading. A related kind of bias arises when the elementary regions in the data set do not have the same size. The range of goods contributing to the observed trade flows strongly depends on the size of each elementary region, leading to a spatial version of selection bias.

The existence of the threshold arises from the interplay between the economic structure of the model and its spatial properties. It is not something that two-, three-, or four-location cases would reveal. The value of the threshold is closely tied to the dimensionality¹⁰ of the spatial configuration. If the spatial geometry is roughly one-dimensional, meaning that economic agents are arranged along a line or a circle, the threshold lies at one particular value for the elasticity of substitution. If economic

¹⁰ The value of the threshold is a linear function of the dimension of space. It is meaningful to consider zero-dimensional cases as well. This corresponds to spatial configurations with just a few (point-like) locations. Here the threshold condition translates into the requirement that the elasticity of substitution be equal to 1. In this case the utility function becomes Cobb-Douglas, which is known to exhibit behavior qualitatively different from the cases with elasticity of substitution greater than 1.

agents are spread through a two-dimensional geometry, the value of the threshold is significantly higher.

The solution method used here is easy to generalize to more complex situations. For example, even though the focus of this work is on static models, dynamic models can be solved in a similar fashion. Adding uncertainty does not represent an obstacle, nor does the addition of differentiated goods with different elasticities of substitution.

The present work is related to two overlapping strands of economic¹¹ research. The first one is concerned with various aspects of empirical data on trade flows (which are generally consistent with the ‘gravity model of trade’). The analysis here is most closely connected to the four models of Krugman (1980), Eaton and Kortum (2002), Melitz (2003), and Anderson and van Wincoop (2003)/Armington (1969), each associated with empirical literature¹² too rich to explicitly cite here.

The other strand of related research studies the influence of international borders on trade flows (McCallum (1995), Anderson and van Wincoop (2003), Behrens, Er-tur, and Koch (2007), Rossi-Hansberg (2005)) and on price fluctuations (Engel and Rogers (1996), Gorodnichenko and Tesar (2009), Gopinath, Gourinchas, Hsieh, and Li (forthcoming)).

The rest of the chapter is organized as follows. The next section justifies the use of functional analysis in later parts of the chapter by discussing various pitfalls associated with oversimplified approaches to multi-location economies. Section 1.3 reviews the basics of the representative example of choice, namely the Krugman (1980) model.

¹¹ There is also a close link to the physics literature; see Section 1.9 and Appendix A.11.

¹² Recent examples include Helpman, Melitz, and Yeaple (2004) and Helpman, Melitz, and Rubinstein (2008). In the context of the present work, it is worth noting that Alvarez and Lucas (2007) establish important properties of the Eaton and Kortum (2002) model and provide a basis for solving the model numerically. In addition, they solve the model analytically under the assumption of zero trade costs and under the assumption of ‘symmetric trade costs’ mentioned earlier.

It also introduces certain concepts needed to characterize the comparative statics of the model. Section 1.4 provides a formal (first-order) solution to the model in the form of an infinite series. Section 1.5 uses Fourier series expansion to derive an explicit general solution to the model in the case of a circular geometry. The resulting formula is then used to analyze two special cases: the impact of changes in border costs in Section 1.6, and changes in productivity in Section 1.7. Spherical geometry is discussed in Section 1.8, with spherical harmonic expansion playing the role of the Fourier series expansion. Section 1.9 considers the structure of higher-order terms.

1.2 *Challenges of multi-location models*

Differentiated goods models, as well as a certain type of Ricardian models, typically lead to large non-linear systems of equations.¹³ The number of equations as well as the number of unknowns is proportional to the number of locations considered. It is clearly desirable to be able to theoretically analyze the solutions to these models even when there is a large number of locations. However, with realistic¹⁴ trade costs this represents a technical challenge. Even after (log)-linearization the behavior of the system is far from obvious. The equations become linear,¹⁵ which certainly is a simplification, but the number of equations and unknowns is not reduced. To solve

¹³ An example may be found in Section 1.3, eq. (1.3), where each equation links the GDP at a particular location to the GDP elsewhere in the economy. This particular case corresponds to Krugman (1980), but analogous equations for other models have a very similar structure.

¹⁴ The term ‘realistic trade costs’ here refers to trade costs that increase with distance, but not as fast as to make the per-unit-distance cost also increasing in distance, as discussed in the introduction.

¹⁵ For trade models where already the exact equations are linear, see Baldwin, Forslid, Martin, Ottaviano, and Robert-Nicoud (2005). An example is the ‘footloose capital’ model of Martin and Rogers (1995).

the system, one needs to invert a large matrix, which is an obstacle¹⁶ for the analytic approach.

The present work uses methods of functional analysis to overcome this difficulty. The reader may ask whether it is really necessary to go through all the calculations in order to get a correct picture of the economic phenomena. Could it be that certain shortcuts lead to qualitatively correct results? The rest of the section is devoted to two such possibilities: working with a few locations only (Subsection 1.2.1) and neglecting indirect general equilibrium interdependencies (Subsection 1.2.2).

1.2.1 Working with only a few locations

Let us look at a very simple situation in which economic activity takes place at many different locations. In this example, the physical space is a continuous circle parameterized by the angle $\theta \in (-\pi, \pi]$. At every point, there are profit-maximizing firms, each producing a different variety of differentiated goods. Only local inputs are used in production. Consumer preferences for the varieties correspond to a constant elasticity of substitution $\sigma \in (1, \infty)$. Apart from the monopoly power of the firms, all markets are competitive (and free of any distortions). Both the setup of the model and the equilibrium involve a complete symmetry between different locations on the circle. To have a concrete model in mind, one can consider, for example, the model of Krugman (1980) or Anderson and van Wincoop (2003) / Armington (1969).

Trade costs are of the ‘iceberg’ type and are characterized by the function¹⁷ $\tau(d) = (1 + \alpha d)^\rho$. When any good is transported over a distance d , a fraction $(\tau(d) - 1) / \tau(d)$

¹⁶ Cramer’s rule, which expresses the solution to a linear system of equations in terms of a ratio of determinants, is of little help here. The determinants are so complicated that they provide little insight into the nature of the solution.

¹⁷ The qualitative conclusions of this subsection apply to any trade cost function $\tau(d)$ that is ‘asymptotically power-law,’ in the sense of Subsection 1.4.3.

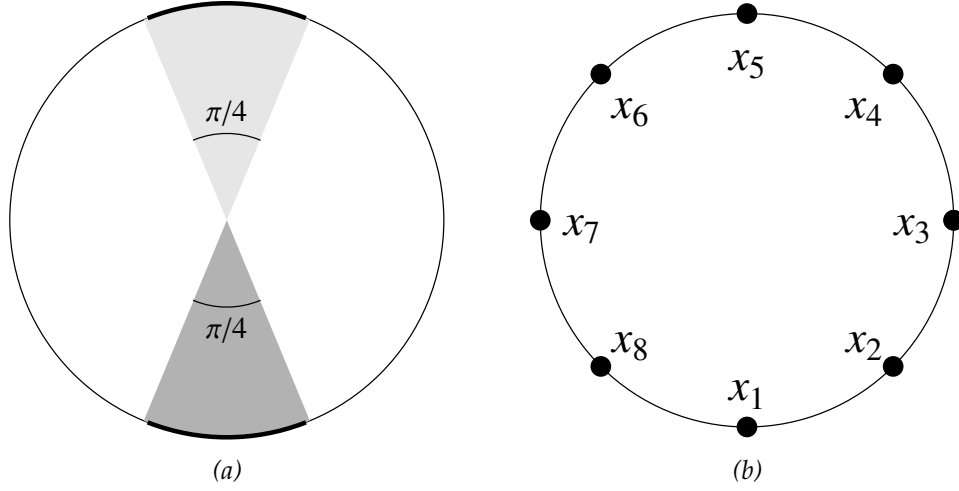


Figure 1.1: (a) The continuous spatial configuration and (b) its discrete approximation.

will be lost. Distance is measured along the circle, and is proportional to the angle between the two locations. The parameters α and ρ are positive exogenously-given constants.

These assumptions are enough to determine the share of expenditures a consumer at location θ spends on products from any given region. For concreteness, consider the consumer located in the middle of the lower shaded angle in Fig. 1.1a and calculate the share s of expenditures on goods from the upper shaded angle in the figure. A short calculation reveals that

$$s = \frac{\pi\alpha R - 1}{1 + \pi\alpha R - (1 + \pi\alpha R)^{\rho(\sigma-1)}} \left(1 - \frac{7}{8} \left(\frac{1 + \pi\alpha R}{1 + \frac{7}{8}\pi\alpha R} \right)^{\rho(\sigma-1)} \right),$$

where R is the radius of the circle. In the large-radius limit, the expression for s simplifies.

$$\lim_{R \rightarrow \infty} s = \begin{cases} 1 - \left(\frac{7}{8}\right)^{1-\rho(\sigma-1)} & \text{for } \rho(\sigma-1) < 1, \\ 0 & \text{for } \rho(\sigma-1) > 1. \end{cases} \quad (1.1)$$

Now suppose that we approximate the circle with a small and fixed number of locations, say eight, as in Fig. 1.1b. If the radius of the circle is very large, consumers at x_1 find varieties produced at other locations very expensive relative to those from x_1 . They will spend almost all of their income on local products. As a result, the counterpart¹⁸ of s approaches zero as $R \rightarrow \infty$ even when $\rho(\sigma-1) < 1$.

This line of reasoning leads to the conclusion that it is impossible to qualitatively reproduce the correct result (1.1) with a finite and fixed number of locations.¹⁹ It is worth emphasizing that the word ‘fixed’ is important in the last sentence. The behavior of the continuous model may be reproduced with a discrete one. To do that, one has to increase the number of locations properly with the radius of the circle when taking the large-space limit. In other words, there is nothing special about working with a continuum of locations from the beginning. What is responsible for the failure of the few-location model is not the discreteness of space, but the fact that additional locations are not added when the radius of the circle is increased.

¹⁸ In the discrete approximation, there is just one location, namely x_5 , at the position of the upper shaded angle of the continuous case. For this reason, the discrete counterpart of s is the share of expenditures of consumers at x_1 on products from x_5 .

¹⁹ The reader may ask whether it is possible to make the few-location model correctly reproduce the qualitative behavior of the continuous model by a simple modification of its assumptions. What if we assume that even goods produced and consumed at the same location have to travel a certain distance, say one-half of the spacing between neighboring locations? It turns out that such assumption does not lead to the desired outcome. It is true that for $\rho(\sigma-1) < 1$ the counterpart of s will be non-zero in the large-space limit. However, under the same assumption the limit of the counterpart of s remains large even in the case $\rho(\sigma-1) > 1$. Moreover, the magnitude of the deviation from (1.1) depends strongly on the arbitrary choice of the number of locations in the discrete model. The departure from the correct value is attenuated only if the number of locations is chosen to be large, contradicting the purpose of the approximation.

1.2.2 *Neglecting changes in general equilibrium effects*

We have seen that one simple way of avoiding algebraic complications, namely working with only a few locations, leads to an impasse. Another way to circumvent the difficulty is to neglect general equilibrium feedback effects when performing comparative statics exercises. In principle, such approach could yield qualitatively correct results. It turns out, however, that even the signs of the resulting quantities may be incorrect, as discussed in Appendix A.1.

To answer the questions raised in the introduction, it is necessary to work with a model involving many locations and to incorporate all general equilibrium effects.

1.3 *The Krugman model*

1.3.1 *Production and transportation*

Consider the static model²⁰ of trade described in Krugman (1980). The spatial geometry consists of N locations x_i with $i = 1, 2, \dots, N$. There is a single factor of production, referred to as labor. Labor markets are competitive, and labor is inelastically supplied. Its endowment at location x_i will be denoted $L_{(x_i)}$. There is a continuum of varieties of goods, each produced by a different monopolistically competitive firm at a single location. Individual varieties are labeled by $\omega \in \Omega$, where Ω is the variety space. To produce an amount q of all varieties between ω and $\omega + d\omega$, for some infinitesimal measure $d\omega$ of varieties, the firms need $F d\omega$ units of labor to cover their fixed overhead costs, and additional $q d\omega$ units of labor to cover their variable costs. Note that this choice corresponds to a particular normalization of the measure of quantity of the goods.

²⁰ The introductory exposition closely follows that of Eaton and Kortum (in progress). The reader may consult this reference for more detail on the derivation of the main equations of the model.

The model uses the ‘iceberg’ specification of trade costs. The goods can be transported from any location x_i to any location x_j , but a fraction $(\tau_{(x_i, x_j)} - 1) / \tau_{(x_i, x_j)}$ will be lost on the way, making the total marginal cost $\tau_{(x_i, x_j)}$ times higher than the manufacturing marginal cost. For obvious reasons, $\tau_{(x_i, x_j)} \geq 1$.

Entry into the industry is free. Consequently, the firms earn zero profits. Given this assumption, the reader can easily verify that if the elasticity of substitution between any two varieties is σ , the firm will find it optimal to spend $\sigma - 1$ times more on variable costs than on fixed costs. As a result, the total measure of varieties produced at x_i is $H_{(x_i)} = \frac{1}{\sigma F} L_{(x_i)}$ in this case.

1.3.2 Consumption

The per-capita consumer utility at a particular location is given by

$$u = \left(\int q^{\frac{\sigma-1}{\sigma}}(\omega) d\omega \right)^{\frac{\sigma}{\sigma-1}},$$

where $q(\omega)$ represents the per capita consumption of variety ω , $\sigma > 1$ is the elasticity of substitution, and the integral is over all varieties available. The per capita spending $p(\omega) q(\omega)$ on variety ω is given by

$$p(\omega) q(\omega) = \left(\frac{p(\omega)}{P} \right)^{1-\sigma} c.$$

Here $p(\omega)$ denotes the price of variety ω , the per capita consumption expenditure is $c = \int p(\omega) q(\omega) d\omega$, and the local price index P is defined as

$$P = \left(\int p^{1-\sigma}(\omega) d\omega \right)^{\frac{1}{1-\sigma}}.$$

To avoid terminological complications, each person is endowed with one unit of labor, and per capita and per unit labor quantities coincide. GDP per capita will be denoted y , to be consistent with the notation for consumption per capita.

1.3.3 Closing the model

The GDP²¹ $y_{(x_i)} L_{(x_i)}$ at location x_i is equal to the revenue its firms collect from the measure $\frac{1}{\sigma F} L_{(x_i)}$ of varieties they produce,

$$y_{(x_i)} = \frac{1}{\sigma F} \sum_{j=1}^N \left(\frac{p_{(x_i, x_j)}}{P_{(x_j)}} \right)^{1-\sigma} c_{(x_j)} L_{(x_j)}.$$

Here $p_{(x_i, x_j)}$ is the price firms from x_i charge at x_j . Setting the markup over total marginal cost $p_{(x_i, x_j)} / (\tau_{(x_i, x_j)} y_{(x_i)})$ to its optimal value of $\sigma / (\sigma - 1)$ and imposing budget constraints $y_{(x_j)} = c_{(x_j)}$, the equation becomes

$$y_{(x_i)} = \frac{1}{\sigma F} \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} \sum_{j=1}^N \left(\frac{y_{(x_i)} \tau_{(x_i, x_j)}}{P_{(x_j)}} \right)^{1-\sigma} y_{(x_j)} L_{(x_j)},$$

with the price index given as

$$P_{(x_j)} = \frac{\sigma}{\sigma - 1} \left(\frac{1}{\sigma F} \sum_k \tau_{(x_k, x_j)}^{1-\sigma} y_{(x_k)}^{1-\sigma} L_{(x_k)} \right)^{\frac{1}{1-\sigma}}. \quad (1.2)$$

Combining the last two equations yields

$$y_{(x_i)}^\sigma = \sum_{j=1}^N \frac{\tau_{(x_i, x_j)}^{1-\sigma} y_{(x_j)} L_{(x_j)}}{\sum_{k=1}^N \tau_{(x_k, x_j)}^{1-\sigma} y_{(x_k)}^{1-\sigma} L_{(x_k)}}. \quad (1.3)$$

²¹ Note that local wages are equal to the local GDP per capita, because labor is the only factor of production and firms earn zero profits.

This is a set of N equations that must hold in equilibrium, and together they determine the economic outcome. The choice of units in which y is measured is arbitrary.²² We are free to pick a numéraire good and normalize its price to 1. (In the subsequent discussion, a different, more abstract condition will be imposed, in order to keep the calculations simple.)

1.3.4 Comparative statics - part 1

The rest of the section discusses the comparative statics of the Krugman model, motivates the definition of the GDP propagator, and establishes its basic properties. Readers interested primarily in the concrete results, not in their detailed derivation, may proceed to Section 1.4.

Consider a small change in trade costs,²³ with the goal of evaluating the induced change in GDP at different places. For ease of notation, denote $T_{(x_i, x_j)} \equiv \tau_{(x_i, x_j)}^{1-\sigma}$. This quantity is sometimes referred to as freeness of trade. The GDP equations are

$$y_{(x_i)} = \left(\sum_{j=1}^N \frac{T_{(x_i, x_j)} y_{(x_j)} L_{(x_j)}}{\sum_{k=1}^N T_{(x_k, x_j)} y_{(x_k)}^{1-\sigma} L_{(x_k)}} \right)^{\frac{1}{\sigma}}. \quad (1.4)$$

Suppose we know y corresponding to some particular T . We are interested in the change $y \rightarrow y + dy$ caused by a change $T \rightarrow T + dT$. Here $y \equiv (y_{(x_1)}, \dots, y_{(x_N)})'$ and T is a collection of $T_{(x_i, x_j)}$. The standard prescription for deriving first-order comparative statics is to differentiate both sides of the equation, leading to

²² The set of equations (1.3) is homogeneous in y .

²³ The general method employed in this work is elucidated using simple examples in Appendix A.2.

$$dy_{(x_i)} = \sum_{i=1}^N G_{(x_i, x_j)} L_{(x_j)} dy_{(x_j)} + \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N H_{(x_i, x_j, x_k)} dT_{(x_j, x_k)}. \quad (1.5)$$

Here $L_{(x_j)} G_{(x_i, x_j)}$ is the derivative of the right-hand side of the i th equation (1.4) with respect to $y_{(x_j)}$, and $H_{(x_i, x_j, x_k)}$ is its derivative with respect to $T_{(x_j, x_k)}$. In matrix notation, the set of equations above becomes

$$(1 - GL_N) dy = \sum_{j=1}^N \sum_{k=1}^N H_{(x_j, x_k)} dT_{(x_j, x_k)}, \quad (1.6)$$

with the $N \times N$ matrix G containing elements $G_{(x_i, x_j)}$, and with the N -dimensional vectors $H_{(x_j, x_k)} \equiv \left(H_{(x_1, x_j, x_k)}, \dots, H_{(x_N, x_j, x_k)} \right)'$. The diagonal $N \times N$ matrix $L_N \equiv \text{diag}(L_{(x_1)}, \dots, L_{(x_N)})$ contains the labor endowments of individual locations on the diagonal. The elements of all of these objects can be computed explicitly if y is known.

The next standard step is to use these equations to express dy in terms of $dT_{(x_j, x_k)}$. To achieve that, one may be tempted to multiply both sides of (1.6) by $(1 - GL_N)^{-1}$, but the situation requires more caution because such matrix is not well-defined. The homogeneity of eq. (1.3) implies²⁴ that GL_N has one eigenvalue equal to 1, associated with the eigenvector y : $GL_N y = y$. Consequently, $1 - GL_N$ has a vanishing eigenvalue and cannot be inverted. For this reason, let us pause here to discuss other properties of the matrix G , which will enable us to complete the calculation.

²⁴ If eq. (1.4) is satisfied for some vector y , it must also be satisfied for γy , where γ is a positive number. Replacing y by γy in (1.4), differentiating with respect to γ , and setting $\gamma = 1$ leads to the conclusion that $GL_N y = y$.

1.3.5 The GDP propagator

Performing the differentiation of the right-hand side of (1.4), $G_{(x_i, x_j)}$ can be written as a sum of two parts,

$$G_{(x_i, x_j)} = G_{c, (x_i, x_j)} + G_{p, (x_i, x_j)}, \quad (1.7)$$

with

$$\begin{aligned} G_{c, (x_i, x_j)} &\equiv \frac{1}{\sigma} y_{(x_i)}^{1-\sigma} \frac{T_{(x_i, x_j)}}{\sum_{k=1}^N T_{(x_k, x_j)} y_{(x_k)}^{1-\sigma} L_{(x_k)}}, \\ G_{p, (x_i, x_j)} &\equiv \frac{\sigma - 1}{\sigma} y_{(x_i)}^{1-\sigma} y_{(x_j)}^{-\sigma} \sum_{l=1}^N \frac{T_{(x_i, x_l)} y_{(x_l)} T_{(x_j, x_l)} L_{(x_l)}}{\left(\sum_{k=1}^N T_{(x_k, x_l)} y_{(x_k)}^{1-\sigma} L_{(x_k)} \right)^2} \\ &= \sigma (\sigma - 1) \frac{1}{y_{(x_j)}} \sum_{l=1}^N G_{c, (x_i, x_l)} G_{c, (x_j, x_l)} y_{(x_l)} L_{(x_l)}. \end{aligned}$$

The matrix G will be referred to as the GDP propagator,²⁵ and G_c and G_p are its ‘consumption part’ and ‘production part’, respectively. These objects capture the strength of GDP spillovers from one location to another.

The intuition behind these expressions is simple. The GDP at location x_j will affect the GDP at x_i through two different channels. The first channel relates to the consumption at x_j , and corresponds to $G_{c, (x_i, x_j)}$. Location x_i is influenced by the consumption at x_j since firms from x_i have customers there. If GDP increases at x_j , the firms will receive more revenue. This is the reason why $G_{c, (x_i, x_j)}$ is positive. The second channel is more closely related to the production at x_j , and is captured by $G_{p, (x_i, x_j)}$. Firms from x_i compete with firms from x_j for customers elsewhere. Higher $y_{(x_j)}$ means more expensive products from x_j , raising the revenue that firms from

²⁵ The algebraic framework used in this chapter is an adaptation of the technique of Feynman diagrams, which has become ubiquitous in physics. The term ‘propagator’ is borrowed from that literature.

x_i receive at x_l . This is again a positive effect, translating into a positive $G_{p,(x_i,x_j)}$. The first effect is direct, so $G_{c,(x_i,x_j)}$ contains $T_{(x_i,x_j)}$. The second effect is indirect, mediated through a third location x_l . For this reason $G_{p,(x_i,x_j)}$ contains $T_{(x_i,x_l)}T_{(x_j,x_l)}$ with l being summed over. The presence of the T s in the denominators is related to the ‘multilateral resistance’ terms in the corresponding ‘gravity equation’, whose importance has been emphasized by Anderson and van Wincoop (2003).

Notice that if trade costs are not symmetric (in the sense that $T_{(x_i,x_j)} \neq T_{(x_j,x_i)}$), then the matrix $G_{(x_i,x_j)}$ will not in general be symmetric. (Even if $y_{(x_i)}$ is the same everywhere and $G_{p,(x_i,x_j)}$ is symmetric as a consequence, the consumption part $G_{c,(x_j,x_i)}$ of the propagator can still be asymmetric.)

The N -dimensional vector space to which y belongs can be thought of as a one-dimensional space spanned by y times an $(N-1)$ -dimensional vector space \hat{Y}_{N-1} whose elements \hat{y} satisfy $y^T L_N \hat{y} = 0$. We already know that the action of GL_N preserves the one dimensional space: $GL_N y = y$. But it is also true²⁶ that the $(N-1)$ -dimensional space \hat{Y}_{N-1} is preserved by the action of this matrix. (In other words, if $y^T L_N \hat{y} = 0$, then also $y^T L_N (GL_N \hat{y}) = 0$.)

Because both the space \hat{Y}_{N-1} and the space spanned by the vector y are preserved by the action of GL_N , the matrix GL_N may be written as

$$GL_N = \mathbf{P}_{span\{y\}} GL_N \mathbf{P}_{span\{y\}} + \mathbf{P}_{\hat{Y}_{N-1}} GL_N \mathbf{P}_{\hat{Y}_{N-1}} = \mathbf{P}_{span\{y\}} + \mathbf{P}_{\hat{Y}_{N-1}} GL_N \mathbf{P}_{\hat{Y}_{N-1}}, \quad (1.8)$$

where $\mathbf{P}_{span\{y\}}$ is the projector onto the one-dimensional space generated by the vector y , and $\mathbf{P}_{\hat{Y}_{N-1}}$ is the projector onto \hat{Y}_{N-1} .

²⁶ To verify this property, it is sufficient to show that $G^T L_N y = ay$ for some constant a . Direct evaluation using the expressions for $G_{c,(x_i,x_j)}$ and $G_{p,(x_i,x_j)}$ above confirms that this is indeed the case with $a = 1$, i.e. that $G^T L_N y = y$.

1.3.6 Comparative statics - part 2

Now let us go back to the discussion of (1.6). We have not imposed any normalization condition on $y + dy$ yet. The international trade literature typically chooses a definite good to serve as numéraire, and normalizes its price to 1. Such choice would be inconvenient in the present context. To take advantage of the decomposition (1.8), we need to impose the more abstract²⁷ condition $y^T L_N dy = 0$, i.e. $dy \in \hat{Y}_{N-1}$. It follows that $GL_N dy \in \hat{Y}_{N-1}$, and as a result of (1.6), also that $\sum_{j=1}^N \sum_{k=1}^N H_{(x_j, x_k)} dT_{(x_j, x_k)} \in \hat{Y}_{N-1}$. Thanks to these properties, the equation (1.6) can be written as

$$\mathbf{P}_{\hat{Y}_{N-1}} (1 - GL_N) \mathbf{P}_{\hat{Y}_{N-1}} dy = \sum_{j=1}^N \sum_{k=1}^N H_{(x_j, x_k)} dT_{(x_j, x_k)}.$$

Since dy and the right-hand side of this equation belong to \hat{Y}_{N-1} , and since the product $\mathbf{P}_{\hat{Y}_{N-1}} (1 - GL_N) \mathbf{P}_{\hat{Y}_{N-1}}$ is an operator on \hat{Y}_{N-1} , we can restrict attention to that space and conclude that²⁸

$$dy = \left(\mathbf{P}_{\hat{Y}_{N-1}} (1 - GL_N) \mathbf{P}_{\hat{Y}_{N-1}} \right)^{-1} \sum_{j=1}^N \sum_{k=1}^N H_{(x_j, x_k)} dT_{(x_j, x_k)}. \quad (1.9)$$

Here, of course, the inversion is performed in \hat{Y}_{N-1} , not in the full N -dimensional space. As discussed in Subsection 1.3.4, GL_N has one eigenvalue equal to 1 and associated with the eigenvector y . Stability of the system implies that all other eigenvalues are smaller than 1 in absolute value. For this reason $\mathbf{P}_{\hat{Y}_{N-1}} (1 - GL_N) \mathbf{P}_{\hat{Y}_{N-1}}$ is invertible in \hat{Y}_{N-1} , and the final expression for dy is well-defined.

²⁷ If all elements of the vector y have the same magnitude, this condition translates into the requirement that the total (nominal) GDP remain fixed as the trade costs change. More generally, the quantity kept fixed is a weighted average of the GDP at individual locations. The same condition may be interpreted in terms of wages, since these are identically equal to the GDP per capita in this model.

²⁸ The continuous-space analog of this equation is the relation (1.21) in Section 1.4.

1.4 The Krugman model in continuous space

While introductory exposition is simpler with a finite number of locations, the examples discussed below will involve continuous space. Retaining a fine discrete grid in the model would not lead to any additional economic insights, and the continuous-space examples provide greater algebraic convenience. The equations of the model may easily be translated into continuum notation.

Let the spatial geometry be a continuous space with points parameterized by a vector of coordinates x . In general, the space can be curved. The coordinates are chosen arbitrarily. Denote the labor element²⁹ at location x by $dL(x)$. The equation (1.3) for GDP becomes

$$y^\sigma(x) = \int \frac{T(x, x') y(x')}{\int T(x'', x') y^{1-\sigma}(x'') dL(x'')} dL(x'), \quad (1.10)$$

where $T(x, x') \equiv \tau^{1-\sigma}(x, x')$. The degree of interdependence between different locations is captured by the GDP propagator³⁰ defined³¹ as

$$G(x, x') = G_c(x, x') + G_p(x, x'), \quad (1.11)$$

²⁹ To follow the discussion, the reader does not have to be familiar with various concepts of differential geometry. Nevertheless, they are useful for expressing $dL(x)$ in more explicit terms. The distances in the physical space are captured by a definite metric tensor whose values depend on x . Denoting its determinant $g(x)$, the endowment of labor $dL(x)$ in a particular coordinate element dx equals $\sqrt{g(x)}dx$ times the labor density.

³⁰ As mentioned in Subsection 1.3.5, the term ‘propagator’ comes from related physics literature.

³¹ For the discrete analog of this definition, see eq. (1.7).

with the ‘consumption part’

$$G_c(x, x') = \frac{1}{\sigma} \frac{y^{1-\sigma}(x) T(x, x')}{\int y^{1-\sigma}(x'') T(x'', x') dL(x'')}, \quad (1.12)$$

and the ‘production part’

$$G_p(x, x') = \sigma(\sigma - 1) \frac{1}{y(x')} \int G_c(x, x'') G_c(x', x'') y(x'') dL(x''). \quad (1.13)$$

Intuitively, the GDP propagator $G(x, x')$ measures how strongly an infinitesimal change in GDP at x' influences the GDP at x . The consumption part reflects the fact that if consumption at x' increases, this will directly increase the sales of firms from x . The production part arises from the fact that increased GDP (wages) at x' make it easier for firms from x to compete in other markets.³²

The GDP propagator satisfies the conditions³³

$$y(x) = \int G(x, x') y(x') dL(x'), \quad y(x) = \int G(x', x) y(x') dL(x'). \quad (1.14)$$

The expression for the price index analogous to (1.2) is now

$$P(x) = \frac{\sigma - 1}{\sigma} \left(\frac{1}{\sigma F} \int T(x', x) y^{1-\sigma}(x') dL(x') \right)^{\frac{1}{1-\sigma}}. \quad (1.15)$$

³² This intuition is discussed in more detail in Subsection 1.3.5.

³³ These are analogous to the conditions $y = GL_N y$ and $y = G^T L_N y$ of Subsection 1.3.5.

1.4.1 Change in the solution in response to a small change in trade costs

Now suppose that the trade costs change³⁴ so that

$$T(x, x') \rightarrow (1 - \kappa b(x, x')) T(x, x'). \quad (1.16)$$

The small but finite parameter κ sets the size of the change, while $b(x, x')$ captures the geometric aspects of the change. For example, if the change under consideration was the introduction of a (proportional) cost of crossing a border, then $b(x, x')$ could be set to one whenever x and x' were separated by the border, and set to zero otherwise.

The GDP equation (1.10) will now take the form

$$y^\sigma(x) = \int \frac{(1 - \kappa b(x, x')) T(x, x') y(x')}{\int (1 - \kappa b(x'', x')) T(x'', x') y^{1-\sigma}(x'') dL(x'')} dL(x'). \quad (1.17)$$

Let us expand the new GDP values in a Taylor series

$$y(x) = y_0(x) + \kappa y_1(x) + \kappa^2 y_2(x) + \dots$$

Here $y_0(x)$ represents the GDP before the change. The functions y_1, y_2, y_3, \dots are required to be orthogonal³⁵ to y_0 , in the sense that $\int y_n(x) y_0(x) dL(x) = 0$ for $n > 0$. These conditions are imposed (instead of fixing the price of a numéraire good) in order to keep the calculations simple. The rationale behind this choice is explained

³⁴ The change in trade costs corresponding to (1.16) is analogous to the change $T \rightarrow T + dT$ considered in the discrete-space case of Subsection 1.3.4. Besides working in continuous space, the difference here is that the change does not have to be infinitesimal.

³⁵ The discrete-space analog of these conditions would be $y_0^T L_N y_n = 0$ for $n > 0$. The space of functions considered here is the space of real square-integrable functions with measure $dL(x)$, i.e. the space of functions f for which $\int f^2(x) dL(x)$ is finite. This space is usually denoted L_2 ; see, for example, Section 15.1 of Stokey, Lucas, and Prescott (1989) for its formal definition. The inner product of functions f and g is defined as $\int f(x) g(x) dL(x)$.

in Subsection 1.3.6.

The main focus of this work is on the first-order change $y_1(x)$. The higher-order terms y_n , $n \geq 2$, may be computed in an analogous way. They are the subject of Section 1.9. An equation for the first-order term $y_1(x)$ can be obtained by plugging the Taylor expansion into the GDP equation and comparing terms of the first-order in κ . The details of the calculation can be found in Appendix A.3. The result is

$$y_1(x) = \int G(x, x') y_1(x') dL(x') + \int B(x, x') y_0(x') dL(x'), \quad (1.18)$$

with the ‘primary impact function’ $B(x, x')$ defined as

$$B(x, x') = -b(x, x') G_c(x, x') + \sigma G_c(x, x') \int b(x'', x') G_c(x'', x') dL(x''). \quad (1.19)$$

Alternatively, using an operator notation, this is

$$y_1(x) = (Gy_1)(x) + (By_0)(x).$$

In general, for a given function $F(x, x')$ the action of the corresponding operator F on a function f will be defined³⁶ as

$$(Ff)(x) = \int F(x, x') f(x') dL(x'). \quad (1.20)$$

Since y_1 is orthogonal to y_0 , and, due to (1.14), so is Gy_1 , it must be that By_0 is

³⁶ Note that the measure $dL(x')$ used here corresponds to the labor endowment. The discrete-space analog would be multiplication by the matrix FL_N .

orthogonal to y_0 as well. The equation for $y_1(x)$ can be iterated indefinitely, giving³⁷

$$y_1(x) = \sum_{n=0}^{\infty} (G^n B y_0)(x). \quad (1.21)$$

Here we used the identity³⁸ $\lim_{n \rightarrow \infty} G^n y_1 = 0$. For later convenience, let us define also the ‘general equilibrium GDP propagator’³⁹ $G_g(x, x')$ as the integral kernel of the operator

$$G_g = - \sum_{n=0}^{\infty} G^{n+1}. \quad (1.22)$$

In terms of G_g , the result (1.21) becomes

$$y_1(x) = ((1 + G_g) B y_0)(x).$$

Another useful expression for y_1 may be obtained using the identity $B = -(1 - \sigma G_c) \tilde{G}_c$, which follows from the definition (1.19) of B . Here \tilde{G}_c is the integral operator corresponding to

$$\tilde{G}_c(x, x') \equiv G_c(x, x') b(x, x'). \quad (1.23)$$

³⁷ The discrete-space counterpart of this equation is the relation (1.9). When interpreting the result (1.21) for y_1 , it is useful to compare it to the expression (A.3) in Appendix A.2, which applies to the case of two endogenous variables. Obviously, the function $B y_0$ plays the role of the vector v . It is an initial effect of the change in κ . Just like in (A.3), this effect has an infinite number of echoes, described by the terms $G^n B y_0$ with positive n .

³⁸ This follows from the fact that any $G^n y_1$ is orthogonal to y_0 , thanks to (1.14), and from the fact that all eigenvalues of G are smaller than 1 in absolute value, except for the one corresponding to the eigenfunction y_0 .

³⁹ This object captures not only the direct interdependencies, but also all the general equilibrium feedback effects.

Denoting also

$$\tilde{g}_c(x) \equiv \int \tilde{G}_c(x, x') dL(x'), \quad (1.24)$$

we have

$$\frac{y_1}{y_0} = - (1 + G_g) (1 - \sigma G_c) \tilde{g}_c. \quad (1.25)$$

For future convenience, let us also introduce the notation

$$\hat{g}_c(x) \equiv \int \tilde{G}_c(x', x) dL(x'). \quad (1.26)$$

Of course, if $G_c(x, x') b(x, x') = G_c(x', x) b(x', x)$ in general, then $\hat{g}_c(x) = \tilde{g}_c(x)$.

The intuition behind the expression (1.19) for the primary impact function $B(x, x')$ is as follows. If a new trade barrier, say a border, is introduced between x and x' , such change is captured by positive $b(x, x')$. There will be two immediate effects on x . First, with the new barrier, firms from x will lose some part of their revenues from x' . This lowers $y(x)$ and is consistent with the first term in (1.19) being negative. Second, it will be easier for these firms to compete with firms from x'' in the market at x' , as long as $b(x'', x')$ is also positive. This effect increases $y(x)$. For this reason, the second term in (1.19) is positive.

1.4.2 Welfare

The welfare of individual agents is characterized by the local GDP per capita adjusted for the local price index, $y^{(P)}(x) \equiv y(x) / P(x)$, where the price index $P(x)$ is given by (1.15) with the replacement $T(x', x) \rightarrow (1 - \kappa b(x', x)) T(x', x)$. Appendix A.4 shows that the price-index-adjusted analog of $y_1(x)$, namely $y_1^{(P)}(x) \equiv \lim_{\kappa \rightarrow 0} (y^{(P)}(x) - y_0^{(P)}(x)) / \kappa$, is given by

$$\frac{y_1^{(P)}(x)}{y_0^{(P)}(x)} = \frac{y_1(x)}{y_0(x)} - \sigma \int G_c(x', x) \frac{y_1(x')}{y_0(x')} dL(x') - \frac{\sigma}{\sigma - 1} \hat{g}_c(x), \quad (1.27)$$

or, in operator notation,

$$\frac{y_1^{(P)}}{y_0^{(P)}} = (1 - \sigma G_c) \frac{y_1}{y_0} - \frac{\sigma}{\sigma - 1} \hat{g}_c. \quad (1.28)$$

Here \hat{g}_c is the function defined in (1.26).

1.4.3 Asymptotically power-law transportation costs

Before specializing to concrete economic situations, let us pause here to clarify the choice of trade cost functions that will be used in the rest of the chapter. The Krugman model uses the ‘iceberg’ form of trade costs, characterized by the quantity $\tau(x, x')$. In principle, the trade costs can depend on many characteristics of location pairs. For example, they are likely to be lower when the two locations share a common language. The present work will abstract from many such possibilities. Instead, the trade costs will take the simple form

$$\tau(x, x') = \tilde{\tau}(d) \tilde{b}(x, x').$$

The first factor $\tilde{\tau}(d)$ corresponds to transportation costs, and depends only on the distance d between x and x' . The second factor $\tilde{b}(x, x')$ represents additional costs, such as the cost of crossing international borders. In baseline cases without any additional trade costs, $\tilde{b}(x, x')$ will be set to 1.

It is common in the empirical literature⁴⁰ to assume that for large d , $\tilde{\tau}(d)$ is well

⁴⁰ See, for example, Anderson and van Wincoop (2003). Note also that the model of Chaney (2011a)

approximated by a power law:

$$\tilde{\tau}(d) \approx (\alpha d)^\rho,$$

with $\rho > 0$ and $\alpha > 0$. Of course, $\tilde{\tau}(d)$ cannot be exactly equal to $(\alpha d)^\rho$ at short distances. Otherwise the obvious restriction $\tau(d) \geq 1$ would be violated⁴¹ for small enough d . There are several convenient functional forms that ensure the $\tau \geq 1$ restriction is satisfied while preserving the power-law behavior at long distances, for example $(1 + \alpha^2 d^2)^{\frac{\rho}{2}}$, $(1 + \alpha d)^\rho$, or $1 + (\alpha d)^\rho$. The present article works with finite geometries, such as a circle or a sphere of radius R . In these cases, closely related functional forms $\left(1 + 4\alpha^2 R^2 \sin^2 \frac{d}{2R}\right)^{\frac{\rho}{2}}$, $\left(1 + 2\alpha R \sin \frac{d}{2R}\right)^\rho$, and $1 + \left(\alpha R \sin \frac{d}{R}\right)^\rho$ provide a greater algebraic convenience. At short distances, these coincide with the previous three, while at long distances they still have the same order of magnitude.

For future purposes, let us mention one important property of the six functional forms above. Define the function $\hat{\tau}(d)$ as

$$\hat{\tau}(d) \equiv \begin{cases} 1 & \text{for } d \leq \frac{1}{\alpha}, \\ (\alpha d)^\rho & \text{for } d \geq \frac{1}{\alpha}. \end{cases}$$

It is true that for each of the six functional forms $\tilde{\tau}(d)$ considered above, there exist⁴²

(with matching frictions similar to the Chaney (2011b) model of trade networks) provides a theoretical justification for the empirical values of the power-law exponent in the international trade gravity equation.

⁴¹ Unless $\tilde{b}(x, x')$ is chosen to precisely compensate for the small magnitude of $\tilde{\tau}(d)$ whenever x and x' are close to each other.

⁴² Concrete values of these coefficients $\{a_l, a_h\}$ that can be used for the functional forms $(1 + \alpha^2 d^2)^{\frac{\rho}{2}}$, $(1 + \alpha d)^\rho$, $1 + (\alpha d)^\rho$, $\left(1 + 4\alpha^2 R^2 \sin^2 \frac{d}{2R}\right)^{\frac{\rho}{2}}$, $\left(1 + 2\alpha R \sin \frac{d}{2R}\right)^\rho$, and $1 + \left(\alpha R \sin \frac{d}{R}\right)^\rho$ are $\{1, 2^{\rho/2}\}$, $\{1, 2^\rho\}$, $\{1, 2\}$, $\{(2/\pi)^\rho, 2^{\rho/2}\}$, $\{(2/\pi)^\rho, 2^\rho\}$, and $\{(2/\pi)^\rho, 2\}$, respectively.

positive constants a_l and a_h independent of R such that

$$a_l \hat{\tau}(d) \leq \tilde{\tau}(d) \leq a_h \hat{\tau}(d) \quad (1.29)$$

for all $d \in [0, \pi R]$. Loosely speaking, this means that these $\tilde{\tau}(d)$ are similar to the simple function $\hat{\tau}(d)$. In general, monotonic functions satisfying this condition will be referred to as ‘asymptotically power-law’, despite the fact that the geometries under consideration have finite R . In the large R limit, the term ‘asymptotically’ regains its conventional meaning.

To simplify notation in the rest of the chapter, the following combination of ρ and σ will be denoted δ :

$$\delta \equiv \frac{1}{2} \rho (\sigma - 1).$$

1.5 The Krugman model on the circle

1.5.1 Basic setup

Consider the case where the spatial geometry is a circle⁴³ of radius R with points parameterized by $\theta \in (-\pi, \pi]$, and where the labor density is constant. Identify the coordinate x with θ . The labor element is now $dL(\theta) = \rho_L d\theta$ with $\rho_L = L / (2\pi)$. The endowment of labor per unit of physical length is ρ_L / R . A baseline solution to the Krugman model corresponding to $\tau(\theta, \theta') = \tilde{\tau}(d)$ is easy to obtain. Due to rotational symmetry, the GDP equation is solved by setting the GDP density to a constant, $y_0(\theta) = y_0$. The GDP propagator $G(\theta, \theta')$ associated with this solution depends only on the distance $d(\theta, \theta')$ between its arguments, defined as the smaller of $|\theta - \theta'|$ and

⁴³ The case of a finite number of locations symmetrically arranged on a circle can be solved in a similar fashion, employing discrete Fourier transform instead of Fourier series expansion.

$2\pi - |\theta - \theta'|$. For this reason, all the information in $G(\theta, \theta')$ can be captured by a function G with only one argument defined by $G(d(\theta, \theta')) = G(\theta, \theta')$. This specifies the single-argument $G(\theta)$ only for $\theta \in [0, \pi]$. For notational convenience, extend it symmetrically to negative arguments, $G(-\theta) = G(\theta)$, and then periodically over the entire real line, $G(\theta + 2\pi n) = G(\theta)$, $n \in \mathbb{Z}$. The action (1.20) of the operator G on any (periodically extended) function $f(\theta)$ on the circle can be written as

$$(Gf)(\theta) = \rho_L (G * f)(\theta) = \rho_L \int_{-\pi}^{\pi} G(\theta - \theta') f(\theta') d\theta'. \quad (1.30)$$

Define also the single argument functions $G_c(\theta)$, $G_p(\theta)$, and $G_g(\theta)$ in a similar way. The symbol $*$ here stands for a 2π -periodic convolution. For any two 2π -periodic functions f and g their 2π -periodic convolution is defined as

$$(f * g)(\theta) = \int_{-\pi}^{\pi} f(\theta - \theta') g(\theta') d\theta'.$$

In the context of the circular geometry, the term ‘convolution’ will always refer to the 2π -periodic convolution.

1.5.2 Expansion in terms of convolution powers of $G_c(\theta)$

We will see that the numerical values of the solutions y_1 can have very different orders of magnitude depending on the values of the parameters of the model, such as ρ or σ . It is desirable to have an intuitive way of finding the correct order of magnitude without performing explicit calculations. For this purpose, let us take a closer look at the mathematical objects the solution contains. Readers interested primarily in the final results for y_1 , not in the properties of individual contributions to it, may proceed to the next subsection.

The formal solution (1.21) can be written as

$$y_1 = \sum_{n=0}^{\infty} \rho_L^n G^{*n} * (By_0), \quad (1.31)$$

where the n th convolution power $G^{*n}(\theta)$ of $G(\theta)$ is the n -fold (2π -periodic) convolution of the function $G(\theta)$ with itself. Because equations (1.11) and (1.13) imply $G(\theta) = G_c(\theta) + \sigma(\sigma - 1)\rho_L G_c^{*2}(\theta)$, the expression for y_1 can be written as

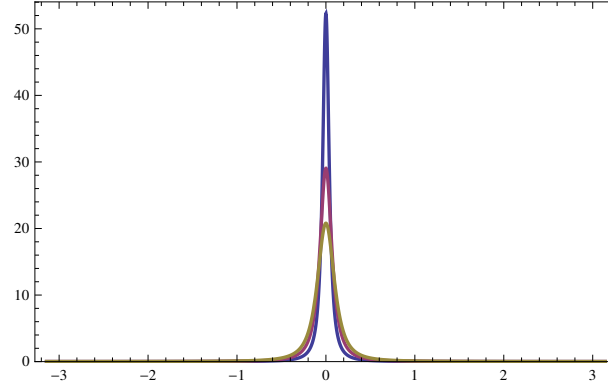
$$y_1 = \sum_{n=0}^{\infty} \rho_L^n (G_c + \sigma(\sigma - 1)\rho_L G_c^{*2})^{*n} * (By_0). \quad (1.32)$$

We see that the right-hand side is a linear combination of various convolution powers G_c^{*m} of the function G_c , convoluted with the function By_0 . In order to gain some intuition about the behavior of y_1 for large R , it is necessary understand what the functions G_c^{*m} look like in that case.

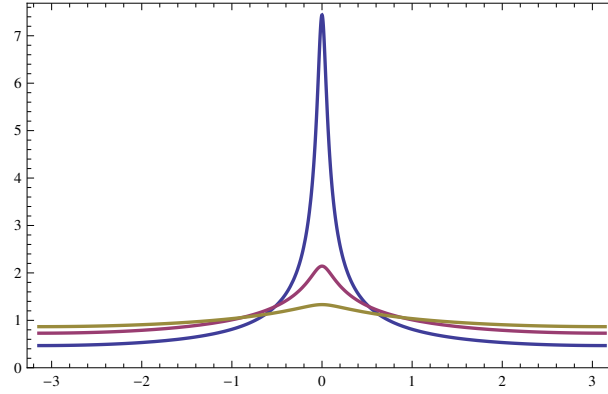
*The large R limit of $G_c^{*2}(\theta)$ and $G_c^{*m}(\theta)$ with $m \geq 3$*

Suppose that R is very large, much larger than $1/\alpha$. The assumption of asymptotically power-law trade costs (1.29) has implications for the behavior of the function $G_c^{*2}(\theta) \equiv \int_{-\pi}^{\pi} G_c(\theta') G_c(\theta - \theta') d\theta'$. A few of its properties are immediately clear. We know that $G_c(\theta)$ is a positive decreasing function of $|\theta| \in [0, \pi]$. As a consequence, the same must be true for $G_c^{*2}(\theta)$. Also, decreasing $\delta \equiv \rho(\sigma - 1)/2$ increases the importance of the tails of the function $G_c(\theta)$, and makes it more spread out. This means that relative to $G_c(\theta)$, any features of the function $G_c^{*2}(\theta)$ will be even more smoothed out. (Note that these observations, as well as those that follow, are consistent with the plots in Fig. 1.2.)

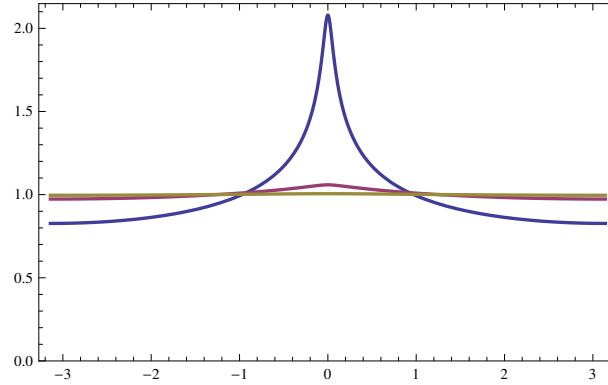
In order to gain a more detailed intuitive understanding of the properties of



(a) $\delta=1.25$



(b) $\delta=0.375$



(c) $\delta=0.125$

Figure 1.2: Plots of $G_c(\theta)$ (highest peak), $G_c^{*2}(\theta)$, and $G_c^{*3}(\theta)$ (lowest peak) for different values of δ . Trade costs are $(1 + 4a^2R^2 \sin^2(\theta/2))^{\rho/2}$ with $\alpha R = 20$. These functions characterize individual contributions to the spreading of economic shocks.

$G_c^{*2}(\theta)$, it is important to know which regions of the integration domain dominate the integral. This issue is technical, and for this reason the derivations are left for Appendix A.5, but the results follow. For $\delta \in (0, \frac{1}{4})$, the main contribution to the integral comes from $|\theta'|$ and $|\theta - \theta'|$ being both of order one. For $\delta \in (\frac{1}{4}, \frac{1}{2})$ it comes from $|\theta'|$ of order $|\theta|$. When $\delta \in (\frac{1}{2}, \infty)$, the integral is dominated by the region where $|\theta'|$ is of order $1/(\alpha R)$ and the region where $|\theta' - \theta|$ is of order $1/(\alpha R)$.

With this knowledge one can make an informed guess about the shape of $G_c^{*2}(\theta)$. With $\delta \in (0, \frac{1}{4})$, the integral is insensitive to what happens at short distances of order $1/(\alpha R)$. For this reason, even though $G_c(\theta)$ has a relatively sharp peak, this feature will be smoothed out in the case of $G_c^{*2}(\theta)$. One can expect the maximum $G_c^{*2}(0)$ to be of the same order of magnitude as the minimum $G_c^{*2}(\pi)$. Moreover, $G_c^{*2}(\pi)$ should have a finite positive limit as $R \rightarrow \infty$.

For $\delta \in (\frac{1}{4}, \frac{1}{2})$, the situation is a little more subtle. For $|\theta|$ of order one, the integral is still dominated by long distances, i.e. by $|\theta'|$ and $|\theta - \theta'|$ of order one. One would therefore expect the minimum $G_c^{*2}(\pi)$ to take similar values as in the previous case. It should stay finite and positive as $R \rightarrow \infty$. By contrast, for small $|\theta|$, say of order $1/(\alpha R)$, the integral is dominated by short distances, i.e. by $|\theta'|$ and $|\theta - \theta'|$ of order $1/(\alpha R)$. This contribution is larger than the contribution of long distances, and as a consequence there should be a substantial peak at $\theta = 0$. In other words, $G_c^{*2}(0) \gg G_c^{*2}(\pi)$.

When $\delta \in (\frac{1}{2}, \infty)$, the story is again relatively simple. Irrespective of the value of θ , the dominant contribution to the integral comes from the short-distance regions, where either $|\theta'|$ or $|\theta' - \theta|$ is of order $1/(\alpha R)$. This means that the shape of the function $G_c^{*2}(\theta)$ should be quite similar to the shape of $G_c(\theta)$, with a large peak and quickly decreasing tails. As $R \rightarrow \infty$, the minimum $G_c^{*2}(\pi)$ should approach

zero. Given the normalization of $G_c(\theta)$, the maximum $G_c^{*2}(0)$ should be of order $G_c(0)/\rho_L$.

Appendix A.5 also contains the derivation of various bounds on the values of $G_c^{*2}(\theta)$. These bounds provide a clearer quantitative picture of the behavior of $G_c^{*2}(\theta)$. As the reader can verify, they are consistent with the intuition just discussed. For the maximum of $G_c^{*2}(\theta)$, which is attained at $\theta = 0$, we have the following bounds

$$\begin{aligned} \frac{1}{2\pi\sigma^2} \frac{(1-2\delta)^2}{1-4\delta} \tilde{a}_l^2 &\lesssim \rho_L^2 G_c^{*2}(0) \lesssim \frac{1}{2\pi\sigma^2} \frac{(1-2\delta)^2}{1-4\delta} \tilde{a}_h^2 && \text{for } \delta \in (0, \tfrac{1}{4}), \\ \frac{(1-2\delta)^2}{2\pi\sigma^2} (\pi\alpha R)^{4\delta-1} \tilde{a}_l^2 &\lesssim \rho_L^2 G_c^{*2}(0) \lesssim \frac{(1-2\delta)^2}{2\pi\sigma^2} (\pi\alpha R)^{4\delta-1} \tilde{a}_h^2 && \text{for } \delta \in (\tfrac{1}{4}, \tfrac{1}{2}), \\ \frac{1}{2\sigma^2} \frac{(2\delta-1)^2}{\delta(4\delta-1)} \alpha R \tilde{a}_l^2 &\lesssim \rho_L^2 G_c^{*2}(0) \lesssim \frac{1}{2\sigma^2} \frac{(2\delta-1)^2}{\delta(4\delta-1)} \alpha R \tilde{a}_h^2 && \text{for } \delta \in (\tfrac{1}{2}, \infty). \end{aligned} \quad (1.33)$$

They are written in terms of the quantity $\rho_L^2 G_c^{*2}(0)$, which does not depend on the choice of units in which labor is measured. The constants $\tilde{a}_l \equiv a_l^{\sigma-1}/a_h^{\sigma-1}$ and $\tilde{a}_h \equiv a_h^{\sigma-1}/a_l^{\sigma-1}$ are defined in terms of the constants appearing in (1.29). The important message these bounds convey is the dependence of $G_c^{*2}(0)$ on the radius R . If $\delta \in (0, \frac{1}{4})$, the peak of $G_c^{*2}(0)$ is relatively small and independent of R . When $\delta \in (\frac{1}{4}, \frac{1}{2})$, the maximum increases as $R^{4\delta-1}$. For $\delta \in (\frac{1}{2}, \infty)$, it increases even faster; it is linearly proportional to R itself.

Now let us look at $G_c^{*2}(0)$ relative to $G_c(0)$.

$$\begin{aligned} \frac{1}{\sigma} \frac{1-2\delta}{1-4\delta} \frac{1}{(\pi\alpha R)^{2\delta}} \tilde{a}_l^2 &\lesssim \frac{\rho_L G_c^{*2}(0)}{G_c(0)} \lesssim \frac{1}{\sigma} \frac{1-2\delta}{1-4\delta} \frac{1}{(\pi\alpha R)^{2\delta}} \tilde{a}_h^2 && \text{for } \delta \in (0, \tfrac{1}{4}), \\ \frac{1}{\sigma} \frac{1-2\delta}{(\pi\alpha R)^{1-2\delta}} \tilde{a}_l^2 &\lesssim \frac{\rho_L G_c^{*2}(0)}{G_c(0)} \lesssim \frac{1}{\sigma} \frac{1-2\delta}{(\pi\alpha R)^{1-2\delta}} \tilde{a}_h^2 && \text{for } \delta \in (\tfrac{1}{4}, \tfrac{1}{2}), \\ \frac{1}{\sigma} \frac{4\delta-2}{4\delta-1} \tilde{a}_l^2 &\lesssim \frac{\rho_L G_c^{*2}(0)}{G_c(0)} \lesssim \frac{1}{\sigma} \frac{4\delta-2}{4\delta-1} \tilde{a}_h^2 && \text{for } \delta \in (\tfrac{1}{2}, \infty). \end{aligned} \quad (1.34)$$

When $\delta \in (0, \frac{1}{2})$, the ratio $G_c^{*2}(0)/G_c(0)$ decreases with R . This means that $G_c^{*2}(0)$ is quite small relative to $G_c(0)$, which is consistent with significant smoothing out. If $\delta \in (\frac{1}{2}, \infty)$, the ratio is independent of R . The peak of $G_c(0)$ is preserved to a large

extent by the convolution.

For the minimum of $G_c^{*2}(\theta)$ at $\theta = \pi$, we have

$$\begin{aligned} \frac{\pi^{4\delta-2}}{4\sigma^2} (1-2\delta)^2 I(\pi) \tilde{a}_l^2 &\lesssim \rho_L^2 G_c^{*2}(\pi) \lesssim \frac{\pi^{4\delta-2}}{4\sigma^2} (1-2\delta)^2 I(\pi) \tilde{a}_h^2 \quad \text{for } \delta \in (0, \tfrac{1}{2}), \\ \frac{1}{2\pi\sigma^2} \frac{2\delta-1}{2\delta} \frac{1}{(\pi\alpha R)^{2\delta-1}} \tilde{a}_l^2 &\lesssim \rho_L^2 G_c^{*2}(\pi) \lesssim \frac{1}{2\pi\sigma^2} \frac{2\delta-1}{2\delta} \frac{1}{(\pi\alpha R)^{2\delta-1}} \tilde{a}_h^2 \quad \text{for } \delta \in (\tfrac{1}{2}, \infty). \end{aligned} \quad (1.35)$$

The function I is defined in (A.9). Its value $I(\pi)$ is independent of R and is roughly of order one when other parameters do not take extreme values. We see that the minimum is independent of R when $\delta \in (0, \frac{1}{2})$, and decreases with R when $\delta \in (\frac{1}{2}, \infty)$.

The last set of inequalities presented here is

$$\begin{aligned} \frac{\pi^{4\delta-2}}{4\sigma^2} (1-2\delta)^2 I(\theta) \tilde{a}_l^2 &\lesssim \rho_L^2 G_c^{*2}(\theta) \lesssim \frac{\pi^{4\delta-2}}{4\sigma^2} (1-2\delta)^2 I(\theta) \tilde{a}_h^2 \quad \text{for } \delta \in (0, \tfrac{1}{2}), \\ \frac{1}{\sigma} \rho_L G_c(\theta) \tilde{a}_l &\lesssim \rho_L^2 G_c^{*2}(\theta) \lesssim \frac{1}{\sigma} \rho_L G_c(\theta) \tilde{a}_h \quad \text{for } \delta \in (\tfrac{1}{2}, \infty). \end{aligned} \quad (1.36)$$

These hold for $|\theta|$ much greater than $1/(\alpha R)$. When $\delta \in (0, \frac{1}{4})$, the function $I(\theta)$ is roughly of order one for any θ . For $\delta \in (\frac{1}{4}, \frac{1}{2})$, it is roughly of order one when $|\theta|$ is of order one. As $|\theta|$ decreases, the function increases indefinitely. But remember that the bound itself is valid only if $|\theta| \gg 1/(\alpha R)$. (A more careful analysis reveals that in this case the peak of $G_c^{*2}(\theta)$ is not very important, it does not contribute much when $G_c^{*2}(\theta)$ is integrated over θ .) When $\delta \in (\frac{1}{2}, \infty)$, the bound implies that $G_c^{*2}(\theta)$ has tails that look similar to those of $G_c(\theta)$.

Here we discussed only $G_c^{*2}(\theta)$, but $G_c^{*m}(\theta)$ with a low $m > 2$ behave in a similar fashion, as the reader can confirm by the same methods. The only qualitative difference is that for $\delta \in (\frac{1}{4}, \frac{1}{2})$ and a high enough m , it ceases to be true that $G_c^{*m}(0) \rightarrow \infty$ as $R \rightarrow \infty$.

1.5.3 General solution for y_1 and $y_1^{(P)}$

The evaluation of $y_1(\theta)$ and $y_1^{(P)}(\theta)$ can be performed using Fourier series. A square integrable function $f(\theta)$ on the circle may be decomposed as

$$f(\theta) = \sum_{n=-\infty}^{\infty} f_n e^{in\theta}, \quad (1.37)$$

where $i = \sqrt{-1}$ is the imaginary unit and the Fourier coefficients f_n are given by

$$f_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta} d\theta. \quad (1.38)$$

In general, the notation used here for the n th Fourier coefficients will be to add subscript n to the symbol of the corresponding function. The convolution theorem for Fourier series states that for two functions f and g the Fourier coefficients $(f * g)_n$ of their (2π -periodic) convolution $f * g$ may be computed by multiplying the Fourier coefficients of the individual functions,

$$(f * g)_n = 2\pi f_n g_n.$$

The operator G acts according to (1.30) as a convolution with $\rho_L G(\theta)$, so

$$(Gf)_n = LG_n f_n.$$

Identical relations hold also for G_c , G_p , and G_g . In the last case, one should remember that G_g is only well defined when acting on functions orthogonal to the constant function y_0 , i.e. on functions f whose zeroth Fourier coefficient f_0 vanishes.

We can now find an expression for the Fourier coefficients $y_{1,n}$ of the function

$y_1(\theta)$. The zeroth coefficient $y_{1,0}$ vanishes since y_1 is chosen to be orthogonal to the constant function y_0 . For nonzero n , applying the convolution theorem to the general expression (1.25) gives

$$\frac{y_{1,n}}{y_0} = - (1 + LG_{g,n}) (1 - \sigma LG_{c,n}) \tilde{g}_{c,n}.$$

This can be further simplified by two identities. The first identity, $LG_{g,n} = 1 / (1 - LG_n) - 1$, comes from the definition (1.22), and the standard formula for the sum of a geometric series. The second identity, $LG_n = (1 + \sigma(\sigma - 1) LG_{c,n}) LG_{c,n}$, is a consequence of (1.11) and (1.13). Together they imply that $-(1 + LG_{g,n}) (1 - \sigma LG_{c,n}) \tilde{g}_{c,n} = -\frac{\tilde{g}_{c,n}}{1 + (\sigma - 1) LG_{c,n}}$. The conclusion is that

$$\frac{y_{1,n}}{y_0} = \begin{cases} 0 & \text{for } n = 0, \\ -\frac{\tilde{g}_{c,n}}{1 + (\sigma - 1) LG_{c,n}} & \text{for } n \neq 0. \end{cases} \quad (1.39)$$

For the local-price-index-adjusted GDP $y_1^{(P)}$ (1.28) leads to

$$\frac{y_{1,n}^{(P)}}{y_0^{(P)}} = \begin{cases} -\frac{\sigma}{\sigma - 1} \hat{g}_{c,0} & \text{for } n = 0, \\ -\frac{1 - \sigma LG_{c,n}}{1 + (\sigma - 1) LG_{c,n}} \tilde{g}_{c,n} - \frac{\sigma}{\sigma - 1} \hat{g}_{c,n} & \text{for } n \neq 0. \end{cases} \quad (1.40)$$

If $b(\theta, \theta') = b(\theta', \theta)$, the Fourier coefficients are real and $\tilde{g}_{c,n} = \hat{g}_{c,n}$. In that case (1.40) simplifies to

$$\frac{y_{1,n}^{(P)}}{y_0^{(P)}} = \begin{cases} -\frac{\sigma}{\sigma - 1} \tilde{g}_{c,0} & \text{for } n = 0, \\ -\frac{2\sigma - 1}{\sigma - 1} \frac{\tilde{g}_{c,n}}{1 + (\sigma - 1) LG_{c,n}} & \text{for } n \neq 0. \end{cases} \quad (1.41)$$

1.5.4 Fourier coefficients of $G_c(\theta)$ for specific functional forms of trade costs

The general formula (1.12) for $G_c(x, x')$ reduces in the case under consideration to

$$G_c(\theta, \theta') = G_c(\theta - \theta') = \frac{1}{\sigma \rho_L} \frac{T(\theta - \theta')}{\int_{-\pi}^{\pi} T(\theta'' - \theta') d\theta''} = \frac{1}{\sigma \rho_L} \frac{T(\theta - \theta')}{\int_{-\pi}^{\pi} T(\theta'') d\theta''}. \quad (1.42)$$

Here, of course, the $T(\theta - \theta')$ corresponds to the trade costs before the introduction of border costs, $T(\theta - \theta') = \tilde{\tau}^{1-\sigma}(\theta - \theta')$. The Fourier coefficients of $G_c(\theta)$ are

$$G_{c,n} = \frac{1}{\sigma L T_0} T_n.$$

Note that this implies that $G_{c,0} = 1/(\sigma L)$, and via (1.11) and (1.13) also that $G_{c,0} = 1/L$, as expected from (1.14).

Subsection (1.4.3) mentioned several convenient functional forms for transportation costs. They all have similar properties. For the purpose of finding analytic solutions to the Krugman model, we will focus mostly on one of them, namely $\tilde{\tau}(d) = \left(1 + 4\alpha^2 R^2 \sin^2 \frac{d}{2R}\right)^{\frac{\rho}{2}}$, but the other ones can be treated similarly. For the functional form of choice, $T(\theta)$ can be written as

$$T(\theta) = \left(\frac{1}{1 + 4\alpha^2 R^2 \sin^2 \frac{\theta}{2}} \right)^{\delta},$$

where the important parameter δ is defined as

$$\delta \equiv \frac{1}{2} \rho (\sigma - 1).$$

An alternative expression for $T(\theta)$ is

$$T(\theta) = Z^{2\delta} \left(\frac{1}{Z^2 \cos^2 \frac{\theta}{2} + \sin^2 \frac{\theta}{2}} \right)^\delta \quad (1.43)$$

with

$$Z^2 \equiv \frac{1}{1 + 4\alpha^2 R^2}$$

As shown in Appendix A.8, the Fourier coefficients of $T(\theta)$ are

$$T_n = \frac{Z^\delta (-1)^n}{(1-\delta)_n} P_{\delta-1}^n \left(\frac{1+Z^2}{2Z} \right).$$

$P_a^b(z)$ is the associated Legendre function.⁴⁴ The Pochhammer symbol $(a)_n$ is defined in terms of the gamma function as $\Gamma(a+n)/\Gamma(a)$, and should not be confused with the notation for Fourier coefficients. For positive integer n , this definition reduces to the n th order polynomial $(a)_n = a(a+1)(a+2)\dots(a+n-1)$. The resulting expression for $G_{c,n}$ is

$$G_{c,n} = \frac{1}{\sigma L} \frac{(-1)^n}{(1-\delta)_n} \frac{P_{\delta-1}^n \left(\frac{1+Z^2}{2Z} \right)}{P_{\delta-1} \left(\frac{1+Z^2}{2Z} \right)}. \quad (1.44)$$

1.6 The impact of border costs

1.6.1 General solution for GDP in the presence of border costs

Now consider the introduction of a small border cost. Let us split the circle into two ‘countries’, country A characterized by $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$ and country B by $\theta \in (-\pi, \frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi]$, separated by a border consisting of two points, $-\frac{\pi}{2}$ and $\frac{\pi}{2}$.

⁴⁴ Mathematica introduces three distinct definitions of associated Legendre functions. The function used here corresponds to the third definition, i.e. to $\text{LegendreP}[\nu, \mu, 3, z]$. See Appendix A.13 for a list of special functions and other mathematical notation.

This assumption is made for simplicity, and generalization to different situations is straightforward. The trade costs are now $\tau(\theta, \theta') = \tilde{\tau}(d) \tilde{b}(\theta, \theta')$, with

$$\tilde{b}(\theta, \theta') \equiv 1 + \tilde{\kappa} 1_{C_A}(\theta) 1_{C_B}(\theta') + \tilde{\kappa} 1_{C_B}(\theta) 1_{C_A}(\theta'),$$

where 1_{C_A} and 1_{C_B} are the country indicator functions. The small positive parameter $\tilde{\kappa}$ is related to the parameter κ considered in the general discussion by $\kappa \equiv 1 - (1 + \tilde{\kappa})^{1-\sigma}$. For small $\tilde{\kappa}$, this is roughly $\kappa \approx (\sigma - 1) \tilde{\kappa}$. In terms of $T(\theta, \theta')$ the change associated with the introduction of the border cost is $T(\theta, \theta') \rightarrow (1 - \kappa b(\theta, \theta')) T(\theta, \theta')$ with

$$b(\theta, \theta') \equiv 1_{C_A}(\theta) 1_{C_B}(\theta') + 1_{C_B}(\theta) 1_{C_A}(\theta').$$

The Fourier coefficients $\tilde{g}_{c,n}$ of the function $\tilde{g}_c(\theta)$ are given in Appendix A.7,

$$\tilde{g}_{c,n} = \begin{cases} 0 & \text{for } n \text{ odd,} \\ \frac{1}{2\sigma} \delta_{0n} - \frac{4(-1)^{\frac{n}{2}}}{\pi^2} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2} & \text{for } n \text{ even.} \end{cases} \quad (1.45)$$

The result (1.39) then becomes

$$\frac{y_{1,n}}{y_0} = \begin{cases} 0 & \text{for } n \text{ odd or zero,} \\ \frac{4(-1)^{\frac{n}{2}}}{\pi^2} \frac{1}{1+(\sigma-1)LG_{c,n}} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2} & \text{for } n \text{ even and nonzero,} \end{cases}$$

while (1.41) gives

$$\frac{y_{1,n}^{(P)}}{y_0} = \begin{cases} -\frac{1}{2} \frac{1}{\sigma-1} + \frac{4}{\pi^2} \frac{\sigma}{\sigma-1} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2} & \text{for } n \text{ zero,} \\ \frac{4(-1)^{\frac{n}{2}}}{\pi^2} \frac{2\sigma-1}{\sigma-1} \frac{1}{1+(\sigma-1)LG_{c,n}} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2} & \text{for } n \text{ even nonzero,} \\ 0 & \text{for } n \text{ odd.} \end{cases} \quad (1.46)$$

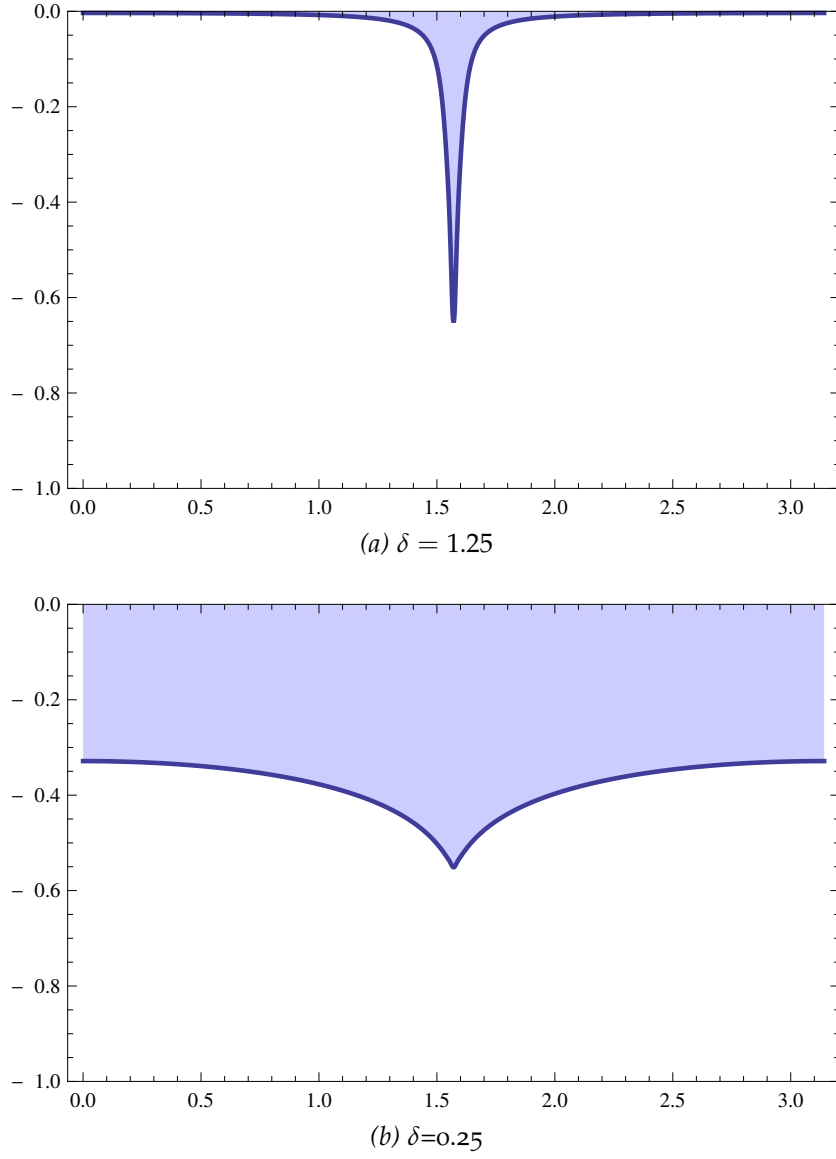


Figure 1.3: Plots of a measure of welfare changes, $(\sigma - 1)y_1^{(P)}(\theta)/y_0^{(P)}$, induced by an increase in the border costs between two semi-circular countries. Only half of each country is shown. Locations on the circle are parameterized by $\theta \in (-\pi, \pi]$. The border is located at $\theta = \pm\pi/2$, and $\theta = 0$ and $\theta = \pi$ correspond to the midpoints of the two countries. In part (a) δ is above the threshold of $1/2$ while in part (b) it is below the threshold. Trade costs are $(1 + 4\alpha^2 R^2 \sin^2(\theta/2))^{\rho/2}$ with $\rho = 0.5$ and $\alpha R = 20$. In part (a) $\sigma = 6$, and in part (b) $\sigma = 2$.

The resulting function $y_1^{(P)}(\theta)$ is plotted in Fig. 1.3 for different values of the parameter δ .

1.6.2 Bounds on $y_1^{(P)}(0)/y_1^{(P)}(\frac{\pi}{2})$

Using the explicit solution (1.45) for functional forms of the trade costs discussed in Subsection 1.4.3, one can derive simple bounds on the values of y_1 . In particular, (1.44) can be used to show that for $\delta < 1/2$,

$$\lim_{R \rightarrow \infty} \frac{y_1^{(P)}(0)}{y_1^{(P)}(\frac{\pi}{2})} \geq \frac{\sigma - 1}{2\sigma - 1} (1 - 2\delta), \quad (1.47)$$

while for $\delta > 1/2$,

$$\lim_{R \rightarrow \infty} \frac{y_1^{(P)}(0)}{y_1^{(P)}(\frac{\pi}{2})} = 0. \quad (1.48)$$

In other words, there is a sharp change of behavior at $\delta = 1/2$ in the large-space limit. Above $1/2$, locations in the middle of the country will not be impacted by the presence of the border at all. Below $1/2$, the impact on the middle of the country will be comparable to that on the border region.

1.7 The impact of changes in productivity

Suppose that the productivity in a particular country changes. How are individual locations inside and outside of this country going to be affected? This question can be answered in a way very similar to the case of the border cost. If the country in question is large, one can consider the same spatial setup as for the border cost. There are two countries, A and B . Suppose that country B , represented by the ‘southern’ semicircle experiences a productivity increase. If productivity in B increases by a factor of $1 + \tilde{\kappa}$, then this is equivalent to decreasing $\tau(x, x')$ from any location x

in country B by the same factor. In terms of the function T , this corresponds to the change $T(x, x') \rightarrow (1 - \kappa b(x, x')) T(x, x')$ with $b(x, x') = -1_{C_B}(x)$ and $\kappa = 1 - (1 + \tilde{\kappa})^{1-\sigma}$. Again, for small $\tilde{\kappa}$, $\kappa \approx (\sigma - 1) \tilde{\kappa}$. Now we can express the main quantities of interest in terms of $G_{c,n}$.

Evaluation of the Fourier coefficients of $\tilde{g}_c(\theta)$ is simple. Since

$$-\tilde{g}_c(\theta) = \rho_L 1_{C_B}(\theta) \int_{-\pi}^{\pi} G_c(\theta - \theta') d\theta' = \frac{1}{\sigma} 1_{C_B}(\theta),$$

they are proportional to the Fourier coefficients (A.14) of the indicator function of country B :

$$\tilde{g}_{c,n} = -\frac{1}{\sigma} 1_{C_B,n} = \begin{cases} -\frac{1}{2\sigma} & \text{for } n = 0, \\ 0 & \text{for } n \text{ even and nonzero,} \\ -\frac{(-1)^{\frac{n+1}{2}}}{\pi\sigma n} & \text{for } n \text{ odd.} \end{cases}$$

Substituting these expressions into (1.39) gives

$$\frac{y_{1,n}}{y_0} = \begin{cases} 0 & \text{for } n \text{ even,} \\ \frac{(-1)^{\frac{n+1}{2}}}{\pi n} \frac{1}{1 + (\sigma - 1) L G_{c,n}} & \text{for } n \text{ odd.} \end{cases}$$

For the local-price-index-adjusted GDP, we should use the general formula (1.40) instead of (1.41), since $b(\theta, \theta')$ is not identically equal to $b(\theta', \theta)$. Remembering that

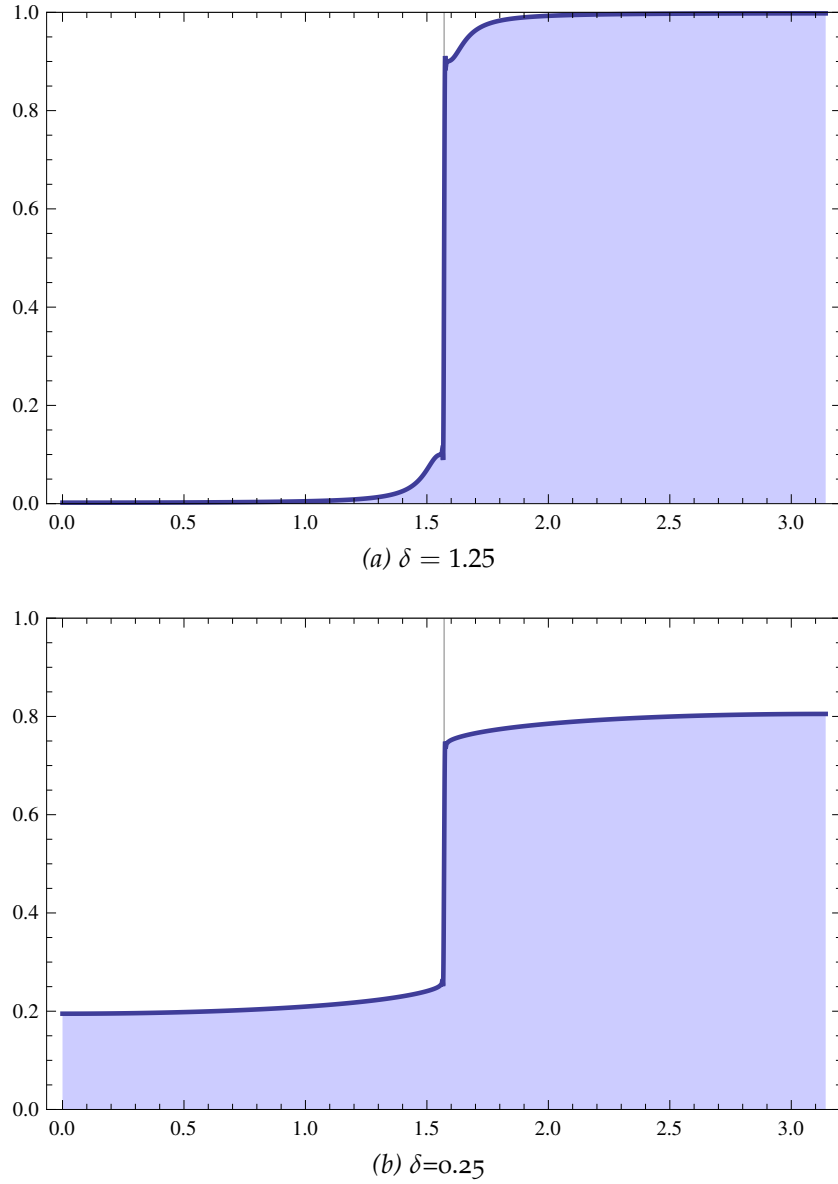


Figure 1.4: This is a productivity change counterpart of Fig. 1.3. It shows a measure of welfare changes, $(\sigma - 1)y_1^{(P)}(\theta)/y_0^{(P)}$, induced by an increase of productivity in country B. Only half of each country is shown. The border is located at $\theta = \pm\pi/2$. As in Fig. 1.3, in part (a) δ is above the threshold of $1/2$ while in part (b) it is below the threshold. Trade costs are $(1 + 4\alpha^2 R^2 \sin^2(\theta/2))^{\rho/2}$ with $\rho = 0.5$ and $\alpha R = 20$. In part (a) $\sigma = 6$, and in part (b) $\sigma = 2$.

$$G_{c,0} = 1/(\sigma L),$$

$$\begin{aligned} -\hat{g}_c(\theta) &= \rho_L \int_{-\pi}^{-\frac{\pi}{2}} G_c(\theta - \theta') d\theta' + \rho_L \int_{\frac{\pi}{2}}^{\pi} G_c(\theta - \theta') d\theta' \\ &= \rho_L \sum_{n=-\infty}^{\infty} G_{c,n} e^{in\theta} \left(\int_{-\pi}^{-\frac{\pi}{2}} e^{-in\theta'} d\theta' + \int_{\frac{\pi}{2}}^{\pi} e^{-in\theta'} d\theta' \right) \\ &= \frac{1}{2\sigma} + \frac{1}{\pi} \sum_{n=-\infty, n \text{ odd}}^{\infty} \frac{(-1)^{\frac{n+1}{2}}}{n} L G_{c,n} e^{in\theta}. \end{aligned}$$

For the individual Fourier coefficients $\hat{g}_{c,n}$ this implies

$$\hat{g}_{c,n} = \begin{cases} -\frac{1}{2\sigma} & \text{for } n = 0, \\ 0 & \text{for } n \text{ even nonzero,} \\ -\frac{(-1)^{\frac{n+1}{2}}}{\pi n} L G_{c,n} & \text{for } n \text{ odd.} \end{cases}$$

The formula (1.40) then yields

$$\begin{aligned} \frac{y_{1,n}^{(P)}}{y_0^{(P)}} &= \begin{cases} -\frac{\sigma}{\sigma-1} \hat{g}_{c,0} & \text{for } n = 0, \\ -\frac{1-\sigma L G_{c,n}}{1+(\sigma-1) L G_{c,n}} \tilde{g}_{c,n} - \frac{\sigma}{\sigma-1} \hat{g}_{c,n} & \text{for } n \neq 0. \end{cases} \\ \frac{y_{1,n}^{(P)}}{y_0^{(P)}} &= \begin{cases} \frac{1}{2} \frac{1}{\sigma-1} & \text{for } n = 0, \\ 0 & \text{for } n \text{ even nonzero,} \\ \frac{(-1)^{\frac{n+1}{2}}}{\pi \sigma n} \left(\frac{1-\sigma L G_{c,n}}{1+(\sigma-1) L G_{c,n}} + \frac{\sigma^2}{\sigma-1} L G_{c,n} \right) & \text{for } n \text{ odd.} \end{cases} \end{aligned}$$

See Fig. 1.4 for plots of $y_1^{(P)}$ for different values of the parameter δ . Again, there is a threshold behavior at $\delta = 1/2$.

1.8 The Krugman model on the sphere

1.8.1 The role of dimensionality

The previous section established that in the large-space limit, the qualitative properties of the Krugman model on the circle with asymptotically power-law trade costs change as $\delta \equiv \rho(\sigma - 1)/2$ crosses the threshold of $1/2$. This value is not universal, however. For spaces of different dimensionality, the value of the threshold is different. In general, for a d_s -dimensional space, the threshold condition is⁴⁵

$$\delta = \frac{d_s}{2}.$$

Clearly, it is of little economic interest to study cases with $d_s \geq 3$. The choice $d_s = 2$, however, is more appropriate for real-world economies than $d_s = 1$.

For this reason, the present section is devoted to the Krugman model on a two-dimensional spatial geometry, the sphere. It turns out that its properties closely resemble the case of the circle, apart from the fact that the role of δ is now played by $\delta/2$.

1.8.2 Basic setup

Let the spatial geometry be a sphere S of radius R parameterized by colatitude $\theta \in [0, \pi]$ and longitude $\varphi \in [0, 2\pi)$. Identify these coordinates with x introduced previously, $x = (\theta, \varphi)$. As in the case of the circle, labor density is chosen to be constant. The labor element is $dL(\theta, \varphi) = \rho_L \sin \theta d\theta d\varphi$ with $\rho_L = L/(4\pi)$. The endowment of labor per unit physical area equals $\rho_L R^2$. Again, the baseline solution

⁴⁵ A hint that this may be the case comes from repeating the calculations that led to eq. (1.1). A careful analysis of geometries of arbitrary dimension provides a confirmation.

corresponds to constant GDP density: $y_0(\theta, \varphi) = y_0$. The GDP propagator $G(x, x')$ depends only on the (rescaled) distance $\tilde{d}(x, x')$ between x and x' given by

$$\cos \tilde{d}(x, x') = \sin \theta \sin \theta' + \cos \theta \cos \theta' \sin(\varphi - \varphi').$$

The information contained in $G(x, x')$ can be captured by a single-argument function G , defined by the relation $G(\tilde{d}(x, x')) = G(x, x')$. The action (1.30) of the operator G can be thought of as a spherical convolution with $\rho_L G(\tilde{d}(x, x'))$,

$$(Gf)(x) = \rho_L (G * f)(x) = \rho_L \int_S G(\tilde{d}(x, x')) f(x') dA(x').$$

Here $dA(x')$ is the (rescaled) area element at point $x' = (\theta', \varphi')$ and may be written as $dA(x') = \sin \theta' d\theta' d\varphi'$. A similar statement holds for G_c, G_p , and G_g and analogously defined functions $G_c(\tilde{d})$, $G_p(\tilde{d})$, and $G_g(\tilde{d})$. Again, it is worth remembering that the action of G_g is defined only on functions orthogonal to the constant function y_0 .

Convolutions on the sphere are a little more subtle than convolutions on the circle. In the case of the circle there is a natural definition of convolution for arbitrary functions as long as the corresponding integral is convergent. On the sphere a natural definition of convolution exists only if at least one of the convolution factors is required to be rotationally symmetric, in the sense that it depends only on θ but not on φ . The functions $G(\tilde{d})$, $G_c(\tilde{d})$, $G_p(\tilde{d})$ and $G_g(\tilde{d})$ all satisfy this requirement, so this is not a source of any difficulty here. The general definition of spherical convolution is

$$(F * f)(x) = \int_S F(\tilde{d}(x, x')) f(x') dA(x'). \quad (1.49)$$

Here F is the function that only depends on θ , identified with \tilde{d} , and f may depend on both spherical coordinates of the point x' .

The spherical analogs of (1.31) and (1.32) take the same form,

$$y_1 = \sum_{n=0}^{\infty} \rho_L^n G^{*n} * (By_0) = \sum_{n=0}^{\infty} \rho_L^n (G_c + \sigma(\sigma-1)\rho_L G_c^{*2})^{*n} * (By_0).$$

The large R results for $G_c^{*2}(\theta)$ (and higher $G_c^{*m}(\theta)$) for the case of the circle have a direct analog here. To avoid repetition, detailed discussion is left for Appendix A.6. As mentioned earlier, the main lesson is that the role of δ (defined as $\rho(\sigma-1)/2$) in the case of the circle is now played by $\delta/2$. Otherwise the qualitative behavior remains the same.

1.8.3 General solution for y_1 and $y_1^{(P)}$

A square integrable function $f(\theta, \varphi)$ on the sphere can be written as

$$f(\theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l f_l^m Y_l^m(\theta, \varphi), \quad (1.50)$$

for some coefficients f_l^m . These coefficients may be computed as

$$f_l^m = \int_S f(\theta, \varphi) Y_l^{m*}(\theta, \varphi) \sqrt{g(x)} dx = \int_0^\pi \int_0^{2\pi} f(\theta, \varphi) Y_l^{m*}(\theta, \varphi) d\varphi \sin \theta d\theta, \quad (1.51)$$

where the star denotes complex conjugation. The spherical harmonic function $Y_l^m(\theta, \varphi)$ of degree l and order m is defined as

$$Y_l^m(\theta, \varphi) = N_l^{|m|} P_l^{|m|}(\cos \theta) e^{im\varphi}.$$

$P_l^{|m|}$ is the associated Legendre polynomial of degree l and order $|m|$, $i = \sqrt{-1}$ is the imaginary unit, and $N_l^{|m|}$ is a positive normalization factor needed to make the system orthonormal (without the Condon-Shortley phase). The general convention

for spherical harmonic coefficients of a function on the sphere is to add the indices l and m to the corresponding symbol of the function. When the index m is zero, it may be omitted. In other words, $f_l \equiv f_l^0$. All spherical harmonics needed here will be of order zero.⁴⁶ They are given more explicitly as

$$Y_l^0(\theta, \varphi) \equiv \frac{\sqrt{2l+1}}{\sqrt{4\pi}} P_l(\cos \theta), \quad (1.52)$$

where P_l is the Legendre polynomial of degree l . According to the convolution theorem on the sphere, spherical harmonic coefficients of the convolution (1.49) are equal to properly normalized products of the spherical harmonic coefficients of the individual convolution factors:

$$(F * f)_l^m = \frac{\sqrt{4\pi}}{\sqrt{2l+1}} F_l f_l^m, \quad (1.53)$$

with $F_l \equiv F_l^0$. For the GDP propagator G this implies

$$(Gf)_l^m = \frac{1}{\sqrt{4\pi}} \frac{1}{\sqrt{2l+1}} L G_l^0 f_l^m,$$

and similarly for G_c, G_p and G_g . Following the same steps as in the case of the circle, we obtain

$$\frac{(y_1)_l^m}{y_0} = \begin{cases} 0 & \text{for } l = 0 \text{ or } m \neq 0, \\ -\frac{\tilde{g}_{c,l}}{1+(\sigma-1)\frac{1}{\sqrt{4\pi}}\frac{1}{\sqrt{2l+1}} L G_{c,l}} & \text{for } l > 0 \text{ and } m = 0, \end{cases} \quad (1.54)$$

⁴⁶ In other applications of the same framework, working with spherical harmonics of nonzero order may be necessary.

$$\frac{(y_1^{(P)})_l^m}{y_0^{(P)}} = \begin{cases} 0 & \text{for } m \neq 0, \\ -\frac{\sigma}{\sigma-1} \hat{g}_{c,0} & \text{for } l = 0 \text{ and } m = 0, \\ -\frac{\sqrt{4\pi}\sqrt{2l+1}-\sigma LG_{c,l}}{\sqrt{4\pi}\sqrt{2l+1}+(\sigma-1)LG_{c,l}} \tilde{g}_{c,l} - \frac{\sigma}{\sigma-1} \hat{g}_{c,l} & \text{for } l > 0 \text{ and } m = 0. \end{cases} \quad (1.55)$$

1.8.4 Solutions for specific functional forms of trade costs

Proceeding in analogy with the case of the circle,

$$G_c(\theta, \varphi, \theta', \varphi') = G_c(\tilde{d}(\theta, \varphi, \theta', \varphi')) = \frac{1}{\sigma \rho_L} \frac{T(\tilde{d}(\theta, \varphi, \theta', \varphi'))}{\int_0^\pi \int_0^{2\pi} T(\tilde{d}(\theta'', \varphi'', \theta', \varphi')) \sin \theta'' d\varphi'' d\theta''},$$

$$G_c(\theta) = \frac{\sqrt{4\pi}}{\sigma L T_0^0} T(\theta).$$

The spherical harmonic coefficients of $G_c(\theta)$ are

$$(G_c)_l^m = \frac{\sqrt{4\pi}}{\sigma L T_0^0} T_l^m.$$

Note that $G_0^0 = \sqrt{4\pi}/L$, which is consistent with (1.14). For transportation costs of the form $\tilde{\tau}(d) = \left(1 + 4\alpha^2 R^2 \sin^2 \frac{d}{2R}\right)^{\frac{\delta}{2}}$,

$$T(\theta) = Z^{2\delta} \left(\frac{1}{Z^2 \cos^2 \frac{\theta}{2} + \sin^2 \frac{\theta}{2}} \right)^\delta,$$

with $\delta \equiv \rho(\sigma-1)/21$ and $Z^2 \equiv (1 + 4\alpha^2 R^2)^{-1}$. Because of rotational symmetry $T_l^m = 0$ and $(G_c)_l^m = 0$ for $m \neq 0$. As shown in Appendix A.10, the remaining coefficients are

$$T_l = 2\pi \sqrt{2l+1}(\delta)_l \frac{Z^{\frac{1}{2}+\delta}}{\sqrt{1-Z^2}} P_{\delta-\frac{3}{2}}^{-l-\frac{1}{2}} \left(\frac{1+Z^2}{2Z} \right),$$

$$G_{c,l} = \frac{\sqrt{4\pi} (\delta)_l}{\sigma L} \frac{P_{\delta-\frac{3}{2}}^{-l-\frac{1}{2}} \left(\frac{1+Z^2}{2Z} \right)}{P_{\delta-\frac{3}{2}}^{-\frac{1}{2}} \left(\frac{1+Z^2}{2Z} \right)}.$$

1.8.5 The impact of changes in border costs

The spherical counterpart of the two semicircular ‘countries’ are the northern and the southern hemisphere, $C_A = \{(\theta, \varphi) \mid \theta \in [0, \frac{\pi}{2})\}$, and $C_B = \{(\theta, \varphi) \mid \theta \in (\frac{\pi}{2}, \pi]\}$. The corresponding spherical harmonic coefficients of \tilde{g}_c (and \hat{g}_c) are given by equation (A.19) in Appendix A.9. The values of the coefficients can be used in the expressions (1.54) and (A.19) for the change in GDP. The resulting solutions exhibit the threshold behavior at $\delta = 1$.

1.8.6 The impact of changes in productivity

Similarly to the case of the circle, $\tilde{g}_c(\theta, \varphi)$ equals $-1_{C_B}(\theta, \varphi) / \sigma$. As a result, its spherical harmonic coefficients with non-zero m vanish and the others are proportional to (A.18):

$$\tilde{g}_{c,l} = \begin{cases} -\frac{1}{\sigma} \sqrt{\pi} & \text{for } l = 0, \\ 0 & \text{for } l \text{ even and nonzero,} \\ -\frac{1}{\sigma} \sqrt{\pi} \sqrt{2l+1} \frac{(-1)^{\frac{-l-1}{2}}}{2^l} \frac{(l-1)!}{\frac{l-1}{2}! \frac{l+1}{2}!} & \text{for } l \text{ odd.} \end{cases}$$

The formula (1.54) then gives

$$\frac{(y_1)_l^m}{y_0} = \begin{cases} 0 & \text{for } l \text{ even or } m \neq 0, \\ \frac{1}{\sigma} \frac{(-1)^{\frac{-l-1}{2}}}{2^l} \frac{(l-1)!}{\frac{l-1}{2}! \frac{l+1}{2}!} \frac{2\pi(2l+1)}{\sqrt{4\pi} \sqrt{2l+1} + (\sigma-1) L G_{c,l}} & \text{for } l \text{ odd and } m = 0. \end{cases}$$

Recognizing that $\hat{g}_c(\theta) = -\rho_L(G_c * 1_{C_B})(\theta)$ and applying the spherical convolution theorem (1.53) leads to

$$\hat{g}_{c,l} = \begin{cases} -\frac{1}{2\sigma} & \text{for } l = 0, \\ 0 & \text{for } l \text{ even and nonzero,} \\ -\frac{(-1)^{\frac{-l-1}{2}}}{2^{l+1}} \frac{(l-1)!}{\frac{l-1}{2}! \frac{l+1}{2}!} LG_{c,l} & \text{for } l \text{ odd.} \end{cases}$$

Of course, $(\hat{g}_c)_l^m$ with nonzero m vanish. The price-index-adjusted change in GDP (1.55) becomes

$$\frac{(y_1^{(P)})_l^m}{y_0^{(P)}} = \begin{cases} \frac{1}{2} \frac{1}{\sigma-1} & \text{for } l = 0 \text{ and } m = 0, \\ 0 & \text{for } l \text{ even nonzero or } m \neq 0, \\ \frac{1}{\sigma} \frac{(-1)^{\frac{-l-1}{2}}}{2^l} \frac{(l-1)!}{\frac{l-1}{2}! \frac{l+1}{2}!} \left(\sqrt{\pi} \sqrt{2l+1} \frac{\sqrt{4\pi} \sqrt{2l+1} - \sigma LG_{c,l}}{\sqrt{4\pi} \sqrt{2l+1} + (\sigma-1) LG_{c,l}} + \frac{1}{2} \frac{\sigma^2}{\sigma-1} LG_{c,l} \right) & \text{for } l \text{ odd and } m = 0. \end{cases}$$

As in the case of border cost, the qualitative behavior changes at $\delta = 1$.

1.9 Higher-order terms

The first-order changes in GDP $y_1(x)$ capture the full impact of changes in trade costs when κ is very small. There is an analytic way to evaluate this impact even for larger κ . The goal of this section is to provide basic insight into how this can be achieved.

As in Section 1.4, suppose that there is an initial equilibrium with GDP equal to $y_0(x)$, and consider a change in trade costs. If the trade costs were characterized initially by $T(x, x')$ and after the change by $(1 - \kappa b(x, x')) T(x, x')$, the new GDP

equation reads

$$y^\sigma(x) = \int \frac{(1 - \kappa b(x, x')) T(x, x') y(x')}{\int (1 - \kappa b(x'', x')) T(x'', x') y^{1-\sigma}(x'') dL(x'')} dL(x'). \quad (1.56)$$

The Taylor series of the new solution $y(x)$ is

$$y(x) = y_0(x) + \kappa y_1(x) + \kappa^2 y_2(x) + \dots$$

The functions $y_n(x)$ are required to satisfy $\int y_n(x) y_0(x) dL(x) = 0, n > 0$. In Section 1.4 we saw that inserting the Taylor expansion into (1.56) and comparing terms linear in κ provides an equation that determines y_1 in terms of y_0 . An analogous statement can be made for the higher-order terms as well. Denote

$$\Delta y(x) \equiv y(x) - y_0(x) = \kappa y_1(x) + \kappa^2 y_2(x) + \dots$$

The GDP equation (1.56) can be expanded as⁴⁷

$$\begin{aligned} & \Delta y(x) - (G\Delta y)(x) \\ &= - (G\Delta y)(x) + y_0(x) \sum_{m=1}^{\infty} \frac{1}{m!} \left(\frac{1}{\sigma}\right)_m y_0^{-m\sigma}(x) \\ & \{ \int T(x, x') y_0(x') \sum_{k=1}^{\infty} \frac{(-1)^k}{(N(x'))^{k+1}} dL(x') \times \\ & (\int (1 - \kappa b(x'', x')) T(x'', x') y_0^{1-\sigma}(x'') \sum_{j=0}^{\infty} \frac{(1-\sigma)_j}{j!} y_0^{-j}(x'') (\Delta y(x''))^j dL(x'') \\ & - N(x')^k \} \\ & + \int T(x, x') (\Delta y(x') - \kappa b(x, x') y_0(x') - \kappa b(x, x') \Delta y(x')) \sum_{k=0}^{\infty} \frac{(-1)^k}{(N(x'))^{k+1}} dL(x') \times \\ & (\int (1 - \kappa b(x'', x')) T(x'', x') y_0^{1-\sigma}(x'') \sum_{j=0}^{\infty} \frac{(1-\sigma)_j}{j!} y_0^{-j}(x'') (\Delta y(x''))^j dL(x'') \\ & - N(x')^k \}^m, \end{aligned} \quad (1.57)$$

⁴⁷ Note that the two different summations over k have different starting points.

where $N(x) \equiv \int y_0^{1-\sigma}(x') T(x', x) dL(x')$ and $(a)_n$ is the Pochhammer symbol. The terms $-(G\Delta y)(x)$ which are present on both sides would cancel, of course. However, when the equation is written in this way, it has an important property. On the left-hand side the term proportional to $\kappa^n, n > 0$ is simply $\kappa^n y_n(x) - \kappa^n (Gy_n)(x)$. On the right-hand side, the term proportional to κ^n contains functions y_0, y_1, \dots, y_{n-1} , but does not contain $y_n, y_{n+1}, y_{n+2}, \dots$. This makes the equation useful: one can first solve for y_1 (as in the previous sections), then for y_2, y_3 , etc. To be more precise, at each step the equation allows one to compute only the function $y_n - Gy_n$ directly. To recover y_n itself, one can use the identity

$$y_n(x) = ((1 + G_g)(y_n - Gy_n))(x).$$

The method⁴⁸ just described provides a way to express any y_n in terms of y_0, T , and b . The resulting expressions may seem very complicated, but they are not. The value of y_n can be written as a sum of a finite number of expressions. These can be evaluated explicitly by the same techniques that were used to compute the first-order terms. Derivation of individual equations from (1.57), as well as the process of solving them, is greatly simplified by a diagrammatic technique⁴⁹ analogous to the

⁴⁸ The mathematical insights underlying the calculation framework introduced here are most closely associated with Richard Feynman. He observed in 1940s – just like Ernst Stückelberg years earlier – that solutions to certain complicated physics problems can be obtained by evaluating series of terms, and that each one of these terms may be represented by a simple cartoon. These “Feynman diagrams” play two different roles. They ensure that one does not get lost in the algebra, and they provide an intuitive way of thinking about the mechanism the model in question represents.

Interestingly, another line of Feynman’s thinking has already influenced other parts of economics; the Feynman-Kac formula is often used in financial economics and related fields. Although this is a mathematically related topic, the typical series of Feynman diagrams with multivalent vertices are not present there. This ultimately follows from the fact that the variables representing physical space here represent state space in the financial application of the Feynman-Kac formula.

⁴⁹ There are two versions of the diagrammatic technique. The first one has important consequences primarily for y_n with $n > 1$. The second one is slightly more complicated, but provides insight into the structure of y_1 .

method of Feynman diagrams in physics.

1.10 Conclusion

Traditional models of international and intranational trade, as well as models introduced in the last decade, have some unexpected spatial properties. As we have seen, under standard assumptions used in the empirical literature, their behavior is highly sensitive to the precise values of their parameters. Naturally, such high sensitivity can lead to strong biases in various estimation procedures. This raises the question: to what extent are existing empirical results affected by such biases?

To address this issue, future empirical work can employ trade models based on familiar principles, but rich enough to include economic sectors with heterogeneous characteristics. The present work provides a convenient way to study the properties of these models analytically, without having to rely on individual numerical solutions, each generated for a single point in a large parameter space. The results of such analytic inquiry will lead to more appropriate model selection for empirical estimation.

2. PASS-THROUGH AS AN ECONOMIC TOOL

(W)e may prepare the way for using, as we go, illustrations drawn from the incidence of taxation to throw side-lights on the problem of value. For indeed a great part of economic science is occupied with the diffusion throughout the community of economic changes which primarily affect some particular branch of production or consumption; and there is scarcely any economic principle which cannot be aptly illustrated by a discussion of the shifting of the effects of some tax...

–Alfred Marshall, “Principles of Economics”, Book V, Chapter IX

2.1 *Introduction*

Following Marshall (1890), standard treatments of a range of topics in perfectly competitive markets are typically taught and analyzed in relationship to tax incidence. For example, Chetty (2009) surveys recent work in public finance that builds on the fact that incidence is often a “sufficient statistic” for various welfare analysis to reduce the number of structural assumptions needed to reach welfare conclusions. Virtually all of this work, however, assumes perfect competition, while much of contemporary economic analysis assumes firms have market power. In this paper we show how the principle of using incidence as an analytic tool, sufficient statistic and key economic parameter extends to imperfectly competitive models, unifying a num-

ber of previous disparate literatures. We survey existing and establish new results relating local and global incidence of various sorts to one another and to properties of the demand and cost system. We then apply these to the analysis of a range of economic applications ranging from platform markets to merger analysis.

We consider a very general model of symmetric imperfect competition, nesting Cournot, differentiated Bertrand, monopolistic competition, conjectural variations and other models. Imperfect competition alters the logic of incidence in two key ways. First, as we discuss in Section 2.2 and has been understood at least since the work of Cournot (1838), the rate $\rho \equiv \frac{dp}{dt}$ at which a tax on a firm is *passed-through* to consumer prices depends on the curvature of consumer demand as well as on the elasticities of supply and demand that determine it in a perfectly competitive market. Second, as less widely discussed in the literature¹ and considered in Section 2.3, tax burden is not simply “split” between the two sides because market power induces a pre-existing distortion. Thus the total magnitude of incidence, and not just its split, is determined by the pass-through rate and the degree of competitiveness.

As a result, pass-through determines the incidence of shocks among consumer surplus, producer and excess burden. Subsection 2.3.1 argues, sufficient taxes eliminate the market, and therefore both consumer surplus and profits, entirely so integrating up pass-through we show that the global division of surplus is given by a simple, level-of-competition-contingent transformation of the average pass-through. We then show that, when there are constant returns to scale, the effect of an exogenous increase in supply on quantity supplied to the market in equilibrium is a simple transformation of the pass-through rate. The incidence of such exogenous supply

¹ For example, while Turner (2012) explicitly uses the pass-through formula for monopoly and cites an earlier version of this paper, she assumes the tax burden is simply split according to the pass-through rate.

shifts is also determined by a simple pass-through-based formula. Because sufficient exogenous increases in supply eliminate deadweight loss and monopoly profits, the global deadweight loss to profit ratio is also determined by simple transformations of average pass-through. In the case of monopoly, both the consumer surplus-to-profit and deadweight loss-to-profit ratios are exactly averages of the pass-through rate itself with no transformation. Section 2.4 provides a graphical illustration of this particularly sharp result for the case of monopoly or a cartel.

The second half of the paper shows how these incidence principles offer an elegant frame for analyzing a range of classical and more recent models and policy questions related to imperfect competition. While we discuss some well-known existing results that may be re-interpreted in light of the incidence logic, our primary focus is on deriving novel results using these ideas and extending existing results to broader settings. We begin with two, more detailed application in Sections 2.5 and 2.6 and then turn to a variety of other applications that we treat more superficially in Section 2.7.

Section 2.5 considers platform industries, where consumers' utility from participation depends on how many other consumers participate. Optimal policy in such markets typically involves trade-offs between utility gained by consumers from lower prices and that lost by a decline in externalities accruing to inframarginal consumers. Both depend crucially on incidence under many natural parameterizations of preferences and thus simple properties of incidence answer a range of normative questions in two canonical models: the Rochet and Tirole (2003) model of two-sided markets and a natural parameterization of the Becker (1991) model with unidimensional diner heterogeneity.

Section 2.6 considers a range of settings in which a planner must procure the cre-

ation of a new market that will be imperfectly competitive ex-post. Examples include auctions for monopolistic concessions, intellectual property incentives for innovation as considered by Weyl and Tirole (2012) and settings where firms may select among proposals for market structures to propose to a planner as in Armstrong and Vickers (2010) and Nocke and Whinston (2011). Mechanism design in all of these settings has to strike a balance between incorporating the private information of firms and the fact that private profits do not coincide with social welfare. Incidence and its heterogeneity (or manipulability) quantify the degree of this conflict and thus are crucial determinants of optimal design.

Subsection 2.7.1 uses incidence to show that when pass-through is constant, there are constant returns to scale and a merger is small, the “diversion ratio” typically used in policy to evaluate the impact of a merger in a differentiated products industry quantifies the fraction of consumer surplus destroyed by the merger. Pass-through plays a broader role in the analysis of mergers and entry, as we illustrate in Subsection 2.7.2 by making more precise the analysis of Mankiw and Whinston (1986) of entry into homogeneous goods industries. In Subsection 2.7.3, summarize and extend to imperfect competition the work of Aguirre, Cowan, and Vickers (2010) and Cowan (2010) who show that simple incidence properties determine the impacts of third-degree price on output, consumer surplus and social welfare. In Subsection 2.7.4, we show how incidence allows an elegant analysis of nonparametric versions of the question of Spence (1975) and Johnson and Myatt (2006) about how firms should design their products to extract maximal value. In Subsection 2.7.5, we show how pass-through determines the equilibrium prices of various levels of imperfectly competitive supply chains and thus the comparison between profits at various levels. In Subsection 2.7.6, we show how incidence provides a unified framework for analyz-

ing (prominent special cases of) the Farrell (2008) model of welfare in aftermarkets and the contrasting Dixit and Norman (1978) and Becker and Murphy (1993) models of the welfare effects of advertising, while extending the results of all of these. Subsection 2.7.7 shows how incidence illuminates the similarities and differences between terms-of-trade-based models of international trade policy with (Brander and Spencer, 1981; Ossa, 2011) and without (Johnson, 1953–1954) market power. In Subsection 2.7.8, we discuss our past and future work showing that incidence properties are restricted in important ways by many common demand forms and proposing tractable demand forms that avoid these restrictions. Finally, in Subsection 2.7.9, we discuss an empirical application of our framework by Atkin and Donaldson (2012) to the competition and the division of surplus in geographically disparate markets in the developing work.

We conclude in Section 2.8 by discussing some directions future research on related to incidence under imperfect competition might take. Extensions of our results on supply shocks and deadweight loss being the constant-returns-to-scale setting appear in an appendix that follows the main text of the paper.

2.2 *Pass-through*

To simplify the analysis, we focus on symmetric industries. However, we allow for (symmetric) product differentiation and arbitrary conduct.² There are n firms and a representative firm i sells quantity q^i at price p^i and has cost function $C(q^i)$ with associated marginal cost $MC(q^i) = C'(q^i)$. We assume there are no fixed costs and

² We neglect the possibility that goods are complements, rather than substitutes, here. However, a symmetric treatment of the case of perfect or imperfect complements is possible and available on request, exploiting the Sonnenschein (1968) duality between competition and complements.

thus $C(0) = 0$.³ We also assume all functions are twice continuously differentiable. Because we consider only symmetric equilibria, we drop the index i when denoting equilibrium quantities. We denote total demand in the market at symmetric prices as $Q(p)$ and inverse demand when aggregate (symmetric) quantity is Q as $p(Q)$. More generally, we use lower case q to denote individual firm level variables and upper case for industry aggregates; by symmetry in equilibrium, $q = \frac{Q}{n}$.

If firms act as price takers they set $p - MC = 0$ while if they collude, or if $n = 1$ and the firm is a profit-maximizing monopolist, they set $p - MC = -\frac{Q}{Q'} \equiv \mu$, which we refer to as the *aggregate market power*. We therefore use an index $l \equiv \frac{p-MC}{\mu}$ to denote the competitiveness of the market conduct ranging from 0 for perfect competition to 1 for perfect collusion. It is useful to consider how two canonical models of imperfect competition fit into this frame, both nesting monopoly/collusion as special cases.

2.2.1 Examples: Cournot and differentiated Bertrand competition

In the Cournot model products are homogeneous and firms choose q^i , taking as given all other firms' quantities. Then they maximize $p\left(q^i + \sum_{j \neq i} q^j\right) q^i - C(q^i)$ and thus their first-order condition is $p + p'q^i - MC = 0$ or, when q is symmetric across firms at a symmetric equilibrium,

$$p - MC = -\frac{Q}{nQ'} = \frac{\mu}{n}$$

³ We do this for two reasons. First, the existence of fixed costs would change the interpretation of our results on global incidence without changing their substance and the interpretation possible in their absence is simpler to explain. Second, they are ruled out by our global concavity assumption discussed in Subsection 2.2.2 discussed below, which is not necessary for our analysis as we showed in previous drafts of this paper, but which is convenient to avoid unnecessary technicalities.

and thus $l = \frac{1}{n}$. As usual, perfect competition corresponds to $n \rightarrow \infty$, while monopoly corresponds to $n = 1$.

On the other hand, suppose that firms produce symmetrically differentiated products and compete Bertrand (Nash-in-prices). Let q^i denote the output of firm i and p_{-i} denote the (symmetric) price of all other firms. If all other firms' prices are symmetric, firm i maximizes $p^i q^i (p^i, p^{-i}) - C(q^i (p^i, p^{-i}))$. This implies firms' first-order conditions are

$$q^i + q_1^i p^i - MC(q^i) q_1^i \iff p^i - MC = -\frac{q^i}{q_1^i} = \left(1 + \frac{(n-1)q_2^i}{q_1^i}\right) \mu.$$

By Slutsky symmetry $\frac{\partial q^i}{\partial p^{-i}} = \frac{\partial q^{-i}}{\partial p^i}$ so that $q_2^i = \frac{\partial q^i}{\partial p^i}$. Thus $D \equiv -\frac{(n-1)q_2^i}{q_1^i}$ is the fraction of the sales that firm i losses as a result of an increase in price that are recaptured by firm $-i$. We refer to this as the *aggregate diversion ratio* to other firms. Thus at a symmetric equilibrium we have

$$p - MC = (1 - D)\mu$$

and thus $l = 1 - D$. $D = 1$ corresponds to perfect competition because it implies that, if there is any elasticity to the market as a whole, the elasticity of individual firm demands are infinitely more elastic as the market-exiting elasticity is a fraction 0 of the total elasticity. If $D = 0$ then the industry acts just as if it were cartelized because the products have independent demand (are not substitutable).

2.2.2 Pass-through formula

Thus our formulation captures a wide range of natural models of imperfect competition that are typically considered distinct. It also allows for a wide range of

others, including conjectural variations (Bowley, 1924), supply function equilibrium (Klemperer and Meyer, 1989), differentiated products Cournot competition and monopolistic competition (Chamberlin, 1931). As we will see, many of the details of the models giving rise to aggregate behavior are less important than the equilibrium pass-through and incidence behavior they induce.

We now use this broad formulation of competition to derive a basic comparative static: the rate ρ at which a specific (per-unit) tax imposed uniformly on all producers is passed through to consumers. Note that because the physical incidence of a tax is, as usual, irrelevant to its economic incidence, a subsidy to (or parallel shift in the demand of) consumers will lead to a rise in the pre-subsidy price of $1 - \rho$, a fact we will occasionally invoke below.⁴

Now suppose that a specific tax of size t is imposed and that l may be a function of t and p . We assume through out that globally $1 - \frac{MC'Q'}{n} - l\mu' - l_2\mu > 0$ to ensure global “concavity”/stability of the unique symmetric equilibrium. The marginal costs of the product increase by t and thus equilibrium conditions are

$$p - MC - t = l(t, p)\mu(p).$$

Implicitly differentiating with respect to t and letting $\rho \equiv \frac{dp}{dt}$,

$$\rho - MC' \frac{Q'}{n} \rho - 1 = l_1\mu + l_2\mu\rho + l\mu'\rho \implies \rho = \frac{1 + l_1\mu}{1 - \frac{MC'Q'}{n} - l\mu' - l_2\mu}. \quad (2.1)$$

⁴ Jeremy Bulow and we, several years back, independently arrived at versions of this observation, but the elegant and parsimonious argument for it given here is Jeremy’s. Note that this implies a subsidy may lead to an optimal *reduction* in price when $\rho > 1$, only possible in an imperfectly competitive market. The result generalizes those of Baldenius and Reichelstein (2000) who consider only constant returns to scale, monopoly and a few thresholds for ρ that correspond to common notions of (log-)concavity of demand.

Note that this formula generalizes both the classical results on incidence in Cournot competition (Cournot, 1838) and the recent results of Anderson, de Palma, and Keider (2001) on differentiated products competition, showing why both yield such similar results, as well as the symmetric case of Reny, Wilkie, and Williams (2011)'s analysis of pass-through with conjectural variations. For the most part we do not discuss the subscripted l terms' effects on pass-through. However, it is worth noting that if taxes directly ($l_1 > 0$) or indirectly ($l_2 > 0$) make conduct less competitive then clearly pass-through will rise. However under Cournot competition $l_1 = l_2 = 0$ and under differentiated Bertrand $l_1 = 0$ always and $l_2 = 0$ if the diversion ratio is constant in the price.

2.2.3 Discussion

The other two factors determining pass-through are the perfectly competitive term $\frac{MC'Q'}{n}$ and the term arising from imperfect competition $l\mu'$. To interpret the first term, note that in a competitive market, the horizontally summed marginal cost curves of the firms represent the supply curve. In this context, these horizontally summed curves are (the inverse of) $MC\left(\frac{Q}{n}\right)$. Thus, assuming it is monotone, we can write the perfectly competitive supply function $S(p) = nMC^{-1}(p)$. While this is not the true supply function in the imperfectly competitive markets we consider, it is useful to consider it to facilitate comparability with the behavior of perfectly competitive markets. The elasticity of this function is $\epsilon_S \equiv \frac{np}{QMC'}$. On the other hand the demand function is Q and has elasticity $-\frac{Q'p}{Q}$. Thus

$$\frac{MC'Q'}{n} = -\frac{-\frac{Q'p}{Q}}{\frac{np}{QMC'}} = -\frac{\epsilon_D}{\epsilon_S}.$$

Thus when there are decreasing returns-to-scale, the more elastic is demand and the less elastic supply, the lower pass-through is. When there are increasing returns, the more elastic demand is and the more rapidly economies of scale set in, the higher pass-through is. If an industry has highly elastic demand and has capacity constraints, we should expect a small pass-through, but if capacity constraints are not too strict and demand is relatively inelastic, pass-through will not be too depressed. If an industry has strong economies of scale and demand is highly elastic at the industry level then it will have very high pass-through, etc. Most economists are familiar with both empirically measuring and calibrating such elasticities.

However, in addition to these competitive forces, economists beginning with Cournot (1838) have recognized that the (log-)curvature of demand, represented by μ' , matters when and to the extent that competition is imperfect. As we discuss more extensively with examples in work in progress (Fabinger and Weyl, 2012), μ' measures the curvature of the logarithm of demand because $\log(Q)' = \frac{Q'}{Q} = -\frac{1}{\mu}$ so $\log(Q)'' = \frac{\mu'}{\mu^2}$. Therefore log-concave demand always has $\mu' < 0$ and log-convex $\mu' > 0$; if demand is concave then $\mu' < -1$ while if it is convex $\mu' > -1$. Another way of viewing μ' , proposed by Gabaix, Laibson, Li, Li, Resnick, and de Vries (2010), is to notice that if α is the standard tail index for the demand, viewed as a probability distribution of consumer values, then $\mu' = \frac{1}{\alpha}$. For the generalized Pareto/constant pass-through class of demand functions proposed by Bulow and Pfleiderer (1983), which include linear, exponential and constant elasticity as special cases, $\mu' = -1$ for linear, $\mu' = 0$ for exponential and $\mu' = \frac{1}{\epsilon}$ for constant elasticity. A final way to think of μ' is in relationship to risk-aversion: $-Q'(1 + \mu')$ is the Arrow-Pratt measure of relative risk-aversion of the demand function Q if this function were to be viewed as a utility function.

Economists have typically viewed μ' , compared to ϵ_D and ϵ_S , as variously difficult to estimate empirically and form intuitions about; see, for example, Farrell and Shapiro (2010). This attitude strikes us as overly pessimistic. Suppose, for example, that consumer willingness to pay was proportional to income. Then μ' corresponds to the well-known curvature properties of income distributions in the segment of the population representing the marginal consumer of the product. Such properties were used by Saez (2001) to calibrate models of optimal income taxation. In particular, $\alpha \in [1.5, 3] \implies \mu' \in [\frac{1}{3}, \frac{2}{3}]$ appears to fit well in the upper tail for most countries and a much less convex distribution (log-normal) appears to fit lower and middle-range incomes. Superior goods will stretch the distribution of willingness to pay and thus have a higher μ' than this income calculation suggests; inferior goods will compress it and lead to a more concave distribution.

Of course this income example is very specific, though commonly used: consumers' willingness-to-pay for most products are not simply proportional to income. However, identifying other specific sources of heterogeneity in willingness-to-pay may help with calibration. In general, idiosyncratic, niche products with highly dispersed values to different populations will have high μ' while utilitarian products that save a constant amount of money to all will have compressed value distributions and thus low and likely negative μ' .

The relative importance of the competitive factors and of demand curvature is determined by l , the conduct parameter. In a highly competitive market (l close to 0), curvature will not play an important role in determining ρ . In a very uncompetitive market, it may well play a dominant role, especially when the elasticity of supply is high relative to the elasticity of demand so that the competitive factors are not important. Note that in a competitive market $MC' \geq 0$, as otherwise one firm would grow

until the market were no longer competitive. Thus pass-through may not be greater than 1 in a competitive market, whereas it may be greater than 1 when competition is imperfect, though it may also be below 1. Empirical work on pass-through, which we survey in Weyl and Fabinger (2009) and Fabinger and Weyl (2012), indicates that it may take on a range of values including both of these. However, it has not focused thus far on confirming or refuting patterns inspired by theory for its value in different settings.

2.3 Local and global incidence

From here on, we assume away income effects. Thus both consumers and firms have quasi-linear utility and surplus may be computed à la Marshall. In a competitive market, both consumers and firms act as price takers and thus, by Hotelling's Lemma given quasi-linearity, consumers lose for each infinitesimal unit that the tax increases $\frac{dp}{dt}Q = \rho Q$ and producer's lose $(1 - \rho)Q$. Thus letting CS represent consumer surplus and PS producer surplus we have that the *incidence* of a tax $I = \frac{\frac{dCS}{dt}}{\frac{dPS}{dt}} = \frac{\rho}{1-\rho}$. In this case $\rho = \frac{1}{1+\frac{\epsilon_D}{\epsilon_S}}$ and thus we obtain the classic formula

$$I = \frac{\frac{1}{1+\frac{\epsilon_D}{\epsilon_S}}}{1 - \frac{1}{1+\frac{\epsilon_D}{\epsilon_S}}} = \frac{\epsilon_S}{\epsilon_D}.$$

The tax is always born completely by either consumers or the firms and the split is determined by the ratio of elasticities. While Hotelling's Lemma continues to hold for consumers when firms are imperfectly competitive, and thus consumers still lose $-\rho Q$, it does not hold for firms as firms no longer take price as given. Aggregate firm

profits are

$$p(t)Q(p(t)) - tQ(p(t)) - nC\left(\frac{Q(p(t))}{n}\right)$$

and thus the change in profits from an increase in the tax is

$$\rho Q + Q'p\rho - Q - tQ'\rho - MCQ'\rho = Q\left[\rho\left(1 - \frac{p - MC - t}{\mu}\right) - 1\right] = -Q[1 - (1 - l)\rho],$$

where l is defined, analogously to before, as $\frac{p - MC - t}{\mu}$. In the extreme case of monopoly this becomes $-Q$: the monopolist, who optimizes the price, by the envelope theorem only experiences the direct loss from the tax, Q . Thus with imperfect competition incidence is

$$I = \frac{\rho}{1 - (1 - l)\rho}. \quad (2.2)$$

The numerator and denominator of incidence here sum to more than 1: $\rho + 1 - (1 - l)\rho = 1 + l\rho$.⁵ Thus, for every unit of additional tax imposed, $Ql\rho$ is excess burden.

2.3.1 Global incidence

These local incidence calculations can naturally be translated into global calculations of the division of surplus. Increasing a tax sufficiently high, potentially towards infinity, eventually eliminates all gains from trade. We can thus see the gains from trade currently in a market as the result of lowering a tax from this point towards 0. Formally let

$$\bar{t} \equiv \begin{cases} \infty & Q(p) > 0 \forall p \\ \inf_{Q(p)=0} p & \exists p : Q(p) = 0 \end{cases}$$

⁵ The first extension we are aware of the incidence formula to imperfect competition was arrived at simultaneously by us and Atkin and Donaldson (2012), building off conversations based on an earlier version of this paper. Their formulation is slightly less general than ours as it assumes homogeneous products.

Let $p(t)$ be the equilibrium price when the tax is t , $CS(t) \equiv \int_{p(t)}^{\infty} Q(x)dx$, $PS(t) \equiv p(t)Q(p(t)) - nC\left(\frac{Q(p(t))}{n}\right) - tQ(p(t))$. We know that $p(t) \geq t$, as otherwise firms make negative profits, so assuming no fixed costs $CS(\bar{t}) = PS(\bar{t}) = 0$. Thus $CS(0) = -\int_0^{\bar{t}} CS'(t)dt$ and $PS(0) = -\int_0^{\bar{t}} PS'(t)dt$ by the fundamental theorem of calculus. But from our logic above, $CS'(t) = -Q(p(t))\rho(p(t), t)$ while $PS'(t) = -Q(p(t))(1 - [1 - l(p(t), t)]\rho(p(t), t))$. Let $\bar{x} \equiv \frac{\int_{t=0}^{\infty} x(t)Q(p(t))dt}{\int_{t=0}^{\infty} Q(p(t))dt}$. Then, compressing all functions to depend directly only on t and dropping the evaluation at 0, we obtain:

$$\frac{CS}{PS} = \frac{\bar{\rho}}{1 - \bar{\rho} + \bar{l}\bar{\rho}} \quad (2.3)$$

If l is constant as a function of p and t this simplifies to

$$\frac{CS}{PS} = \frac{\bar{\rho}}{1 - (1 - l)\bar{\rho}}. \quad (2.4)$$

This generalizes the bounds on consumer surplus obtained by Anderson and Renault (2003) in three ways. First, it replaces bounds based on global inequalities on μ' with precise average expressions that imply these bounds. Second, it provides much richer results when costs are not constant. Third and most importantly, it applies well beyond Cournot competition, on which those authors focused.⁶ In the case of perfect competition this also provides a global expression for the consumer to producer surplus ratio, based on average elasticities, which we have not been able to find in the literature.

⁶ Our proof is also substantially simpler.

2.3.2 Excess burden

The same approach may be used to provide a global expression for deadweight loss, using the logic discussed above that a tax generates an excess burden of $Ql\rho$. It is thus intuitive that the larger pass-through is the larger is deadweight loss created by the implicit tax of monopoly. To formalize this argument, we focus on the case of constant marginal cost of c ; our appendix extends our results allow more general variable costs.

Consider an industry facing, in addition to a potential tax, exogenous competition in an amount \tilde{Q} ; i.e. rather than demand in the industry being $Q(p)$ it is $Q(p) - \tilde{Q}$. It can be shown, by the same arguments above, that under Cournot and differentiated Bertrand competition the degree of competition is unaffected by the introduction of the exogenous quantity. We omit this derivation for brevity and instead proceed based on a definition of l , valid in these cases, which is analogous to that above:

$$l = -\frac{(p - c - t)Q'}{Q - \tilde{Q}},$$

as μ is now effectively $-\frac{Q-\tilde{Q}}{Q'}$ rather than $-\frac{Q}{Q'}$. To compare the comparative static effect of t and \tilde{Q} , we take the partial derivative of the right hand side with respect to t and \tilde{Q} respectively, ignoring any implicit effects through changes in p . The first, analogously to before, yields $\frac{Q'}{Q-\tilde{Q}}$ and

$$-\frac{Q'(p - c - t)}{(Q - \tilde{Q})^2} = \frac{1}{Q'} l \frac{Q'}{Q - \tilde{Q}}.$$

Thus $\frac{dp}{d\tilde{Q}} = \frac{1}{Q'}l\rho$ or the *quantity pass-through*

$$\rho_q \equiv \frac{dQ}{d\tilde{Q}} = l\rho. \quad (2.5)$$

That is, an increase in exogenous (quantity) competition raises the equilibrium quantity produced by the industry by $l\rho$.

Thus, again together with the degree of competition, pass-through is equivalent to the “strategic” effect of increased quantity when there are constant returns to scale: if total quantity expands by $l\rho$ then the quantity of the industry (not including the exogenous quantity) increases by $l\rho - 1$ and thus there are “strategic substitutes” if $l\rho < 1$ and “strategic complements” if $l\rho > 1$.

Now consider the incidence of this exogenous competition, assuming the exogenous quantity is produced at the same, constant marginal cost that the rest of the industry faces. Industry profits are $[P(Q) - c](Q - \tilde{Q})$. Taking the derivative with respect to \tilde{Q} yields

$$P'l\rho(Q - \tilde{Q}) + (l\rho - 1)(P - c) = M \left(l\rho - 1 - \frac{-(Q - \tilde{Q})P'}{P - c}l\rho \right) = -M[1 + (1 - l)\rho],$$

where $M \equiv P - c$. Note that, now that there is an “external” producer of the good exogenously supplying \tilde{Q} at constant marginal cost c , we must consider the impact on her welfare as well. We add this to consumer welfare and calculate that the impact on consumer-and-external-producer welfare, is $-P'(Q - \tilde{Q})l\rho + M = M(\rho + 1)$. Thus, summing the two terms together, we see that social surplus rises (deadweight loss falls) by $Ml\rho$ when \tilde{Q} increases. Thus the ratio of gained efficiency to lost firm profits is

$$\frac{l\rho}{1 + (1 - l)\rho}.$$

Let Q^{**} be the efficient quantity where $P(Q^{**}) = c$ and Q^* be the equilibrium quantity when $\tilde{Q} = 0$. Then note that if $\tilde{Q} = Q^{**}$ both profits and deadweight loss are 0 as the efficient quantity is supplied and firms have no mark-up and therefore (given constant returns) makes zero profits. Therefore we can apply the same argument that we used to link local and global incidence. In particular letting $\tilde{x} \equiv \frac{\int_{\tilde{q}=0}^{Q^{**}} x(\tilde{Q})M(Q(\tilde{Q}))d\tilde{q}}{\int_{\tilde{q}=0}^{Q^{**}} M(Q(\tilde{Q}))d\tilde{q}}$ and the deadweight loss be defined as $DWL \equiv \int_{Q^*}^{Q^{**}} M(Q)dQ$ we have that,

$$\frac{DWL}{PS} = \frac{\tilde{l}\tilde{\rho}}{1 + \tilde{\rho} - \tilde{l}\tilde{\rho}},$$

or, if l is independent of \tilde{q} and Q ,

$$\frac{DWL}{PS} = \frac{l\tilde{\rho}}{1 + (1 - l)\tilde{\rho}}.$$

This formalizes the intuitive idea that when pass-through is high, excess burden is large relative to profits because consumers bear an important part of the implicit tax created by imperfect competition. It also formalizes the “folk theorem” (Harberger, 1954; Gilbert and Shapiro, 1990) that when competition is closer to perfect, deadweight loss relative to profits is small as a small l makes both the numerator small and the denominator large.

2.4 Graphical illustration

This section provides a simple graphical illustration of our results, focusing on the monopoly case. To facilitate comparison, the mark-up over marginal cost and quantity at the optimum are held constant across all panels in Figure 2.1. In the first row, firms have constant returns technology; we compare two different demand curves

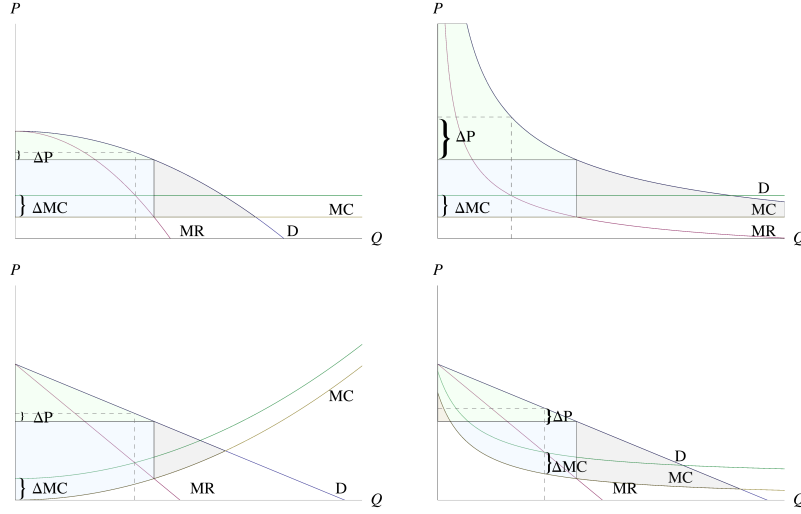


Figure 2.1: Pass-through and the division of surplus for various demand-cost combinations.

Top has $c = 1$; left: $P(q) = 5 - \frac{q^2}{3}$; right: $P(q) = .221 - \frac{1}{Q^2}$. Bottom has $P(q) = \frac{19}{3} - \frac{4}{3}q$; left: $MC(q) = \frac{Q^2}{4}$; right: $MC(q) = \frac{5}{2q+1}$.

with different pass-through rates, given constant returns, drawn from the Bulow and Pfleiderer (1983) constant (linear cost) pass-through demand class. On the left, the concave demand curve has a small pass-through as indicated by the small increase in price induced by increasing the cost and thus moving along the marginal revenue curve and up to the demand curve to the new price. On the right, the convex demand curve has a high pass-through rate. Note that, consistent with our incidence results, consumer surplus and deadweight loss are much larger relative to (the same) profits on the right than on the left.

The second row illustrates a cost-side transformation. Both the left and right panels have the same linear demand and marginal cost at the optimum, but the left panel has an increasing and the right panel a declining marginal cost curve. Pass-through is thus smaller on the left than on the right. While both have the same consumer surplus, monopoly profit is larger than with linear cost on the left, and smaller on the right (in fact, the area above price and below marginal cost on the right should be subtracted from the already-small profits). Thus, again, the ratio of consumer surplus and deadweight loss to profits tracks pass-through.

2.5 *Platforms*

Many, and increasingly many, products (such as credit cards, social networks and operating systems) operated by firms with market power exhibit consumption externalities between users. This has generated a growing literature, surveyed by Rysman (2009), on such *platform* industries. An influential parameterization of such models, proposed by Rochet and Tirole (2003) (RT2003), is for individual to gain a utility from participating proportional to the aggregate number of individuals participating on the platform, but for this constant of proportionality to differ across individuals. Because this model creates trade-offs between the surplus of infra-marginal consumers and prices, incidence plays an important role in its analysis. We therefore consider platforms as our first and most detailed application.

2.5.1 *The Rochet and Tirole (2003) model*

Consider, in particular, the original RT2003 model: a monopoly platform must attract two distinct groups of users, such as holders and acceptors of a credit card or gamers and developers for a video game system. Value is generated by the interac-

tion of these two groups, who are either completely or randomly matched, with the number of users on each of the sides, denote B and S , depending on the *per-interaction* prices p^B and p^S respectively charged. The fractions of users participating at these prices are denoted by $Q^B(p^B)$ and $Q^S(p^S)$ respectively and the total number of interactions is (proportional to) $Q^B(p^B) Q^S(p^S)$. Each Q^I function is a standard demand function: smooth, strictly declining and exhibiting declining marginal revenue. The platform faces a cost c per interaction of providing the service and thus earns profits

$$\pi(p^B, p^S) = (p^B + p^S - c) Q^B(p^B) Q^S(p^S).$$

User surplus on each side, per-interaction, is given by the integral of demand $V^I(p^I) = \int_{p^I}^{\infty} Q^I(x) dx$ while average surplus is $\bar{V}^I(p^I) = \frac{V^I(p^I)}{Q^I(p^I)}$. RT2003, and especially Rochet and Tirole (2006), emphasize the distinction between the *price level* $\bar{p} \equiv p^B + p^S$ and the *price balance*, the division among the two prices, holding fixed the price level.

An elegant feature of the RT2003 model is that the quantity on each side of the market simply multiplies the quantity on the other side, so that given linear per-interaction costs it does not affect optimal pricing. Furthermore the (per-interaction) price on each side of the market enters the optimization on the other side exactly as a subsidy would: the effective marginal cost on side I is $c - p^{-I}$. This simplifies analysis of monopoly pricing: on each side I the monopoly pricing rule is followed taking $c - p^{-I}$ as cost so

$$p^I - (c - p^{-I}) = \mu^I(p^I),$$

where $\mu^I \equiv -\frac{Q^I}{(Q^I)'}$ is the market power term. Noting that $p^I - (c - p^{-I}) = \bar{p} - c$ we

obtain a summary of the first-order conditions as

$$\bar{p} - c = \mu^B(p^B) = \mu^S(p^S). \quad (2.6)$$

Socially optimal pricing, on the other hand, follows Pigou's rule (Weyl, 2009) that price equals marginal cost less positive externalities. Here, the per-interaction externality is the gross willingness of users on the other side to pay for an additional interaction, the sum of their per-interaction price and average surplus. Thus socially optimal pricing requires $p^I = c - p^{-I} - \overline{V^{-I}}(p^{-I})$ or

$$c - \bar{p} = \overline{V^B}(p^B) = \overline{V^S}(p^S). \quad (2.7)$$

It can be shown (Weyl, 2009) that at *any* price level, not merely the optimal ones, the profit-maximizing and consumer surplus optimal balance of prices is determined by the second equation in (2.6) and (2.7) respectively. A profit maximizer seeks to maximize volumes, equating the inverse hazard rates, while a planner maximizes external benefits by equating average surplus.

Let us consider three policy-oriented questions in this context:

1. How do the socially and privately optimal balance of prices compare? Holding fixed the price level, would moving the balance in one direction or the other increase total user welfare?

Consumer surplus-optimal (CSO) balance requires that $\overline{V^B} = \overline{V^S}$ or, equivalently and using $\bar{\rho}$ analogously to Subsection 2.3.1, that $\bar{\rho}^B \mu^B = \bar{\rho}^S \mu^S$ while profit-maximizing balance requires $\mu^B = \mu^S$. Thus the divergence between the two conditions is the factor of $\bar{\rho}$: if $\bar{\rho}^B > \bar{\rho}^S$ the CSO balance calls for lower prices to side *S* than does the profit-maximizing balance as side *B* gains more

infra-marginal surplus from the externalities delivered by side S than vice versa. Thus the comparison of average pass-through on the two sides determines how the balance should be adjusted from its private settings to maximize consumer welfare. RT2003 assume linear demand, which has constant pass-through of $\frac{1}{2}$, on both sides of the market and thus obtain a coincidence between socially and privately optimal price balance.

2. What is the incidence of a tax on interactions (an increase in c)? Is it positive on both sides?

A rise in the price on side $-I$ is equivalent to a cost subsidy on side I ; these are passed through at the pass-through rate to price. Thus

$$\frac{dp^I}{dc} = \rho^I \left(1 - \frac{dp^{-I}}{dc} \right),$$

allowing us to solve out, invoking a symmetric argument, for

$$\frac{dp^I}{dc} = \rho^I \frac{1 - \rho^{-I}}{1 - \rho^B \rho^S}.$$

The second-order conditions for a monopoly optimum require $\rho^B \rho^S < 1$.⁷ As a result, we note that the incidence of side I may be either positive or negative depending on how ρ^{-I} compares to unity; it is possible that it is above unity without violating the second-order condition.⁸ RT2003 thus obtain equal incidence of $\frac{1}{3}$ for each of the firm and consumers on the two sides.

⁷ A small increase in prices on side B in amount ϵ is passed through as a decrease in amount $\rho^S \epsilon$ on side S and this, in turn, would be passed through as an increase of $\rho^B \rho^S \epsilon$ on side B . If this second-round effect is greater, the process is unstable, violating the standard second-order condition.

⁸ Note that observing a shock to cost's effect on the two prices identifies the two pass-through rates.

3. How does an infinitesimally small price control of size δ on one side of the market impact surplus on the other side? In particular, what is the impact of a small price control starting from the monopolist's optimum?

Let us decompose the effect. Total consumer surplus on side I is $\bar{V}^I Q^B Q^S = \bar{\rho}^I \mu^I Q^B Q^S$. The price control on side $-I$ has two effects: it increases the externalities accruing to $-I$ by $-\delta \bar{\rho}^I \mu^I Q^I (Q^{-I})'$. On the other hand it leads to an increase in prices on side I in an amount ρ^I which, by the envelope theorem, costs consumers on side I $\delta \rho^I Q^B Q^S$. The net effect of this is

$$\begin{aligned} -\delta \left(\bar{\rho}^I \mu^I Q^I (Q^{-I})' + \rho^I Q^B Q^S \right) &= \delta Q^B Q^S \left(\bar{\rho}^I \frac{\mu^I}{-\frac{Q^{-I}}{(Q^{-I})'}} - \rho^I \right) \\ &= \delta Q^B Q^S \left(\bar{\rho}^I \frac{\mu^I}{\mu^{-I}} - \rho^I \right), \end{aligned}$$

where the first equation follows just by rearrangement, while the second follows by the definition of μ^{-I} . Note that at the monopoly's pre-control optimum, $\mu^I = \mu^{-I}$ and thus the expression is proportional to $\bar{\rho}^I - \rho^I$. Intuitively, the relative size of the externality compared to pass-through effects should be determined by the divergence between the degree of infra-marginal surplus and the pass-through of the tax arising from the reduction in prices on the other side. Thus whether the other side benefits from or is harmed by a small price control is determined by how average pass-through at prices above the equilibrium compares to its marginal value. This, in turn, is determined by whether pass-through is (on average) increasing, in which case the other side benefits from the control, decreasing, in which case it is harmed, or constant, in which case the two effects are a wash. RT2003 thus obtain that consumers on each side are indifferent to a small decrease in prices on the other side as linear demand

has constant pass-through.

While the RT2003 is obviously very special, it is useful to note that the key special feature that makes incidence so important is not the random matching assumption or two-sidedness but rather the uni-dimensionality of types that creates a simple and known relationship between marginal and infra-marginal consumers' preferences. This feature is shared with the overwhelming majority all platform models we are aware of.⁹

2.5.2 *The Becker (1991) model*

To illustrate this, rather than considering each of these models separately, we instead consider another platform model that differs along as many dimensions as possible, other than the uni-dimensionality of heterogeneity, from the RT2003 model and show how the welfare analysis turns on similar features. Becker (1991) proposes a platform model of restaurant pricing. While his model is much broader, let us consider a natural version of his model with one-dimensional heterogeneity, though of a sort opposite to that in RT2003 and much closer to the other canonical model of two-sided markets by Armstrong (2006). Because of social influences, diners receive a smooth utility of $u(N)$ for visiting the restaurant if a fraction of the population N also wants to go to the restaurant, in addition to an inherent utility from eating there of $\theta - P$ where P is the restaurant's price. Becker assumes that, for exogenous reasons of capacity constraints, the restaurant cannot accommodate more than K individuals and thus earns profits $P \min\{N, K\} - C(\min\{N, K\})$. Nonetheless the firm may find it optimal to set prices low enough to attract $N > K$ diners because raising the price may cause a discontinuous collapse in demand to below K even if θ is drawn from a

⁹ See Rochet and Tirole (2006) and Weyl (2010) for notable exceptions.

continuous distribution as the reduction in demand can cause a downward spiral of diners' social utility falling because other diners defect. In this case we may suppose that K customers of the N interested in a spot are randomly selected.

A natural question is whether a profit-maximizing monopolist will restrict N below what is optimal, as in a standard market. This is not obvious as increasing N increases the amount of rationing, an additional effect not internalized by the firm. A useful way to consider this problem is to imagine the firm can directly control N ; see Weyl (2010) for a justification of how the firm may be able to do this through dynamic pricing. We let $P(N)$ be the prevailing price at which N customers want to go to the restaurant given that N want to go and let f be the continuous distribution function of θ . $P(N)$ is the unique solution for P to $N = \int_{P-u(N)}^{\infty} f(\theta)d\theta$. Let $M(N) \equiv f(P(N) - u(N))$ be the density of marginal diners. By the implicit function theorem $1 = -(P' - u')M$ so $P' = u' - \frac{1}{M}$.

Assuming $N > K$, the interesting case, diner surplus is

$$K \left[\frac{\int_{\theta=P(N)-u(N)}^{\infty} \theta f(\theta)d\theta}{N} + u(N) - P(N) \right].$$

The sign of the monopoly distortion is simply the derivative of this with respect to N as the monopoly does not internalize the diners' surplus while the planner does. Taking this derivative, dividing by the constant K and dropping arguments yields

$$\frac{-(P' - u')MN(P - u) - \int_{\theta=P-u}^{\infty} \theta f(\theta)d\theta}{N^2} + u' - P' = \frac{P - u - \frac{\int_{\theta=P(N)-u(N)}^{\infty} \theta f(\theta)d\theta}{N}}{N} + \frac{1}{M} =$$

$$\frac{P - u - \bar{\theta}}{N} + \frac{1}{M'},$$

where $\bar{\theta} \equiv \frac{\int_{P-u(N)}^{\infty} \theta f(\theta)d\theta}{N}$ which has the same sign as $\frac{N}{M} - \frac{CS}{K} = \mu - \frac{CS}{K}$. By the same

reasoning as above, this is positive if $\bar{\rho} < 1$ and negative if $\bar{\rho} > 1$. Thus if $\bar{\rho} < 1$ there is still a monopoly distortion but if $\bar{\rho} > 1$ a profit-maximizing monopolist would actually attract *too many* diners to the restaurant to wait in line as she fails to internalize the disutility of rationing.¹⁰ Thus we see that incidence plays central role in a range of platform models with unidimensional user heterogeneity with different dimensions of heterogeneity, utility functions, rationing rules, etc.

2.6 *Procuring new markets*

2.6.1 *Concession auctions*

A public authority seeks to select the provider of a monopolistic concession to maximize the social surplus this creates. Suppose that each concession operator will charge a uniform price if she wins the concession, has a single feasible proposal and has private information on both the consumer surplus CS and profits π this proposal will generate. The authority cannot or is unwilling to monitor prices ex-post to avoid monopoly distortions and thus must simply choose the operator generating most surplus.

In addition to purely public settings, similar trade-offs arise when platforms, such as supermarkets (Armstrong and Zhou, 2011) or websites (Edelman, Ostrovsky, and Schwarz, 2007), allow product sellers to display their wares or advertisements for these prominently in exchange for payment, because, as Gomes (2011) argues, the platform has an incentive to internalize the consumer surplus generated by these products in order to profit from consumers on other offerings such as fixed fees for

¹⁰ This logic, and the result it generates, are essentially identical to the analysis of the welfare effects of price controls in competitive markets with perfectly inelastic supply in Bulow and Klemperer (Forthcoming).

using the platform. In these literatures our assumptions of no discrimination, ex-post monitoring or project selection are maintained. In the following two subsections we show how the same principles of incidence will be relevant if these restrictive assumptions are relaxed.

Solving for the optimal mechanism in this multidimensional context is beyond the scope of our analysis here, but some interesting results can be derived directly from the logic of incidence.¹¹ To begin, note that social surplus at the monopoly's optimal price is the sum of consumer surplus and monopoly profits or $(1 + \bar{\rho}) \pi$. Because only π affects the incentives of the various potential operators to seek the concession, the logic of Jehiel and Moldovanu (2001) suggests that it will typically be impossible to use a mechanism to screen for anything other than π . If the planner views $\bar{\rho}$ as symmetrically distributed across firms (conditional on π), then she wishes to select the firm with highest π if and only if $E[(1 + \bar{\rho}) \pi | \pi]$ is ranked in the same way π is. A grossly sufficient condition for this is that $\bar{\rho}$ is distributed independently of π . Clearly if $\bar{\rho}$ is constant across all competitors, this is satisfied, implying Armstrong, Vickers, and Zhou (2009)'s result that if all firms have linear demand and constant returns (which yields constant pass-through of $\frac{1}{2}$) then the ranking of profits and social surplus are identical and a simple auction is optimal.

More generally, an auction should perform reasonably well so long as there is not too strong a negative correlation between $\bar{\rho}$ and π . If such a correlation were too strong, the planner might want to be unresponsive to π because of the implied adverse selection, randomizing among symmetric proposals. Especially in such cases, the authority would seek information that would allow it to handicap the auction

¹¹ For more on two distinctive approaches multidimensional mechanism, see Rochet and Stole (2003) and Veiga and Weyl (2012). The logic of the results given here is closely related to that of the latter paper for obvious reasons.

to favor operators with high expected \bar{p} . Firms with proposals that would involve strict capacity constraints, for example, should be disfavored as this implies rapidly declining returns and thus low pass-through. Firms with proposals that bring valued by an enthusiastic niche of consumers or offer products with high income elasticities should be favored as such products are likely to have highly convex demands and thus a high \bar{p} .¹² We suspect that principles of incidence would also play an important role in the design of the optimal mechanism.

In the above we have focus on the case of ex-post monopoly as this is one of the more common cases for concession auctions. However, if multiple, but a small number of, operators is being selected to supply symmetrically differentiated services and the profits of each industry configuration is observable or can be screened, the same logic as that can be applied, replacing the monopoly incidence expression \bar{p} with the imperfectly competitive incidence $\frac{\bar{p}}{1-(1-l)\bar{p}}$. Now the competitiveness of different arrangements, if this may vary, will also be a desirable screening variable. In the following two sections we focus again on the ex-post monopoly case, with the understanding maintained that the results are extensible to the analogous imperfect competition settings.

2.6.2 Innovation incentives

One restriction in the previous subsection was that the authority could not monitor prices ex-post. In a setting without competition (procuring a single innovation), Weyl and Tirole (2012) consider mechanisms to simultaneously screen for the best project while holding down prices ex-post. Assuming no marginal costs of produc-

¹² We have assumed uniform pricing, but if some firms were able to discriminate more effectively than others, discriminatory firms should be penalized (relative to their willingness to pay) as they will appropriate more consumer surplus.

ing the product ex-post, they consider providing innovators with rewards of the form $T(p^\alpha q^{1-\alpha})$. Higher values of α lead to higher prices, but also raise rewards on a given isoreward curve to innovations with high values of $x \equiv \frac{p}{q}$. In particular, taking a derivative of the rewards

$$T' \cdot [\log(p) - \log(q)] p^\alpha q^{1-\alpha}$$

shows that the marginal reward from raising α for a given innovation is proportional to $\log(x)$ while the social value of an innovation is proportional (on a given isoreward curve) to $(1 + \bar{\rho})x^{1-2\alpha}$. To eliminate the dependence on $\bar{\rho}$, Weyl and Tirole assume innovations are drawn according to a parameterization of demand that ensures $\bar{\rho}$ is constant across all innovations. However, more generally, they show that the value of raising α and thus distortion price ex-post in order to screen more effectively is proportional to the average over all values of k of

$$\text{Cov} \left[\log(x), (1 + \bar{\rho})x^{1-2\alpha} \mid p^\alpha q^{1-\alpha} = k \right].$$

Thus a natural way to enrich their analysis would be to determine, empirically or theoretically, the correlation between $\bar{\rho}$ and various increasing functions of x . If small-market, high-priced goods have more convex demand, as seems plausible, market power will be more attractive than the Weyl and Tirole analysis implies as the above covariance will be greater. If large-market, low-priced goods have more convex demand, on the other hand, market power may be much less useful as a screening device. This could be extended to allow variable costs of producing the innovative product and broader rewards schemes by using properties of incidence more broadly in combination with the more general version of the above covariance formula de-

rived by Veiga and Weyl (2012). Combining this model with ex-ante competition for the right to enter would transform this into a model of ex-post price monitoring in the concession auction model; given the important role of incidence in screening in both models we suspect they would combine in a natural way.

2.6.3 *Project choice/design*

Another restriction we imposed in Subsection 2.6.1 was that each concession operator had a single, exogenous proposal. If, instead, operators can take (potentially costly) actions to affect the characteristics of their proposals or can choose among many potential proposals, moral hazard enters the problem. We consider each of these formulations in turn as they correspond to different strands in the contract theory literature. Again, we assume away competition.

Holmström and Milgrom (1991) consider a model where only one dimension of a multidimensional set of choices by an agent may be observed. In the setting discussed above, π is the natural observable dimension and $\bar{\rho}$ the natural unobservable dimension. Even if not directly observable, profits can be screened by conditioning the probability of granting the agent operation of the concession on payments made, while as discussed above $\bar{\rho}$ is unlikely to be directly screenable. By the logic of Holmström and Milgrom, the more difficult it is to observe $\bar{\rho}$ and the more responsive $\bar{\rho}$ is to effort by the agent, the less incentives should respond to π (the less of the profits generated by effort the agent should be allowed to keep). Designing an optimal mechanism to implement this screening in the absence of ex-post monitoring of profits is beyond our scope here, but the principle is clear.

An alternative formulation, proposed by Armstrong and Vickers (2010) is that the agent may choose one of a finite set of concession projects to propose. The set

available to the agent is unknown to the principal. The principal observes the payoff to herself of a proposal and the payoff to the agent and thus may prohibit certain combinations from being created, but has no access to transfers. They show that the principal should prohibit combinations with high agent payoffs and low principal payoffs, even if these benefit both the principal and agent compared to the status quo, to encourage the agent to choose more socially advantageous proposals, even though this means committing to implementing. They show that this effect grows with the number of proposals are likely to be available and falls with the correlation between the dimensions and with weight is placed on the agent's payoff. In the concession context, the agent payoff is π and the principal's $(1 + \bar{\rho})\pi$. As above, therefore, properties of $\bar{\rho}$ and its correlation with π determine optimal policies.

One particular implication of this style of logic is explored by Nocke and Whinston (2011) who consider proposals to be mergers, many of which could be proposed. They argue that the threshold in terms of consumer welfare gains for a merger to be approved should be higher for mergers between larger firms in a Cournot industry because a greater fraction of the gains from such mergers accrue to the merging partners. That is, the incidence of the efficiency gains from the merger is tilted towards the merger partners. Any other factor that would lead the incidence to be inclined in this direction, such as greater capacity constraints among the firms, more concave joint demand, etc. would lead to an analogous result by the logic of Armstrong and Vickers. Thus a class of principles for merger policy can be derived from incidence logic. In the following section we analyze more closely the connection between industry-wide incidence and the incidence of mergers.

2.7 Other Applications

2.7.1 Mergers

A long line of work (Werden, 1996; Shapiro, 1996; Farrell and Shapiro, 2010) has established a close connection between the impact of mergers in differentiated products industries and the effects of changes in cost. These ideas have been incorporated into policy through the new United States and United Kingdom horizontal merger guidelines. They suggest that agency investigators consider the equivalent of a merger in marginal cost changes to determine its competitive effects. Jaffe and Weyl (2012) show that a (matrix) product of pass-through rates and these equivalent cost changes are a first-order approximation to the effect of a merger on prices, where the approximation's error is proportional to the curvature of the demand system and the square of the size of the equivalent cost changes. Rather than discuss these existing results, in this section we consider the simpler, symmetric, merger-to-monopoly case with constant returns to scale and use the incidence analysis from above to establish more detailed connections between incidence and the effects of mergers. We focus on the case of mergers between the producers of differentiated products and return to the analysis of homogeneous goods in the next subsection on entry.

There are two firms producing symmetrically differentiated products and competing Bertrand. As derived in Subsection 2.2.1, the pre-merger first-order condition for a symmetric equilibrium at common price p is

$$f(p) \equiv p - c - (1 - D(p)) \mu(p) = 0.$$

The post-merger first-order condition is

$$h(p) \equiv p - c - \mu(p) = 0.$$

Following the logic of Jaffe and Weyl, let p_M and p_0 represent the assumed-unique post- and pre-merger prices. Then, assuming h and f are locally invertible,

$$p_M - p_0 = h^{-1}(0) - h^{-1}(h(p_0)) \approx \rho(p_0) D(p_0) \mu(p_0) + O\left(D(p_0)^2 \mu(p_0)^2\right),$$

where ρ is the post-merger pass-through $\frac{1}{1-\mu'}$. Note that because the merger is small ($\rho D \mu|_{p_0}$ is small because D is small) then $\mu(p_0) \approx \mu(p_M)$, $\rho(p_0) \approx \rho(p_M)$ and $Q(p_0) \approx Q(p_M)$ then the price increase is approximately $\rho(p_M)\mu(p_M)D(p_0)$ and the fall in consumer surplus $Q(p_M)\rho(p_M)\mu(p_M)D(p_0)$.

Note that $\mu = p - c$ post-merger so post-merger $PS = \mu(p_M) Q(p_M)$. Thus, by our results of Subsection 2.3.1 consumer surplus post-merger is $\bar{\rho}(p_M) Q(p_M) \mu(p_M)$. Thus if $\bar{\rho}(p_M)$ is close to $\rho(p_M)$, as would be precisely the case if pass-through is constant, we may interpret $D(p_0)$ as approximately *the ratio of the consumer surplus destroyed by the merger to the total post-merger consumer surplus* or $\frac{D(p_0)}{1+D(p_0)}$ as the fraction of pre-merger consumer surplus destroyed by the merger. This makes the diversion ratio a particularly natural gauge of merger impacts.

Similar approximations may be used to derive the effect of mergers on profits ($\rho^2 D^2 \mu$) as in Baker and Bresnahan (1985) or the deadweight loss created by it (for a small merger to monopoly this is the same as the change in consumer surplus as the change in profits is second order by this approximation) using the principle of incidence. We suspect that in the richer, multi-product firm, merger-to-imperfect-competition with nonlinear costs, as analyzed in Jaffe and Weyl (2012), analogous

results may be obtained using the multi-dimensional analogs of the imperfectly competitive incidence formulas from Section 2.3 and bounds based on cost non-linearities as in our appendix.

2.7.2 Entry

Now we consider the effects of entry into an initially n -firm symmetric, Cournot industry with common constant returns production technology. Let q_{n+1} be the equilibrium post-entry, per-firm quantity and q_n be the same before the merger. Entry is equivalent to the creation of an exogenous quantity of q_{n+1} . Let $\hat{\rho}_n^{n+1}$ be the average pass-through taken with respect to changes in the exogenous quantity and not weighted by the mark-up. Then the entry increases industry quantity by $\frac{\hat{\rho}_n^{n+1}}{n}q_{n+1}$ and thus

$$\frac{\hat{\rho}_n^{n+1}}{n}q_{n+1} + nq_n = (n+1)q_{n+1} \implies q_{n+1} = \frac{n}{n+1 - \frac{\hat{\rho}_n^{n+1}}{n}}q_n.$$

Thus output per-firm expand (contracts) if and only if $n < (>)\hat{\rho}_n^{n+1}$. Mankiw and Whinston (1986), ignoring the discreteness of entry, show that entry generates net positive (negative) externalities if other firms' output expands (contracts).¹³ We can make these results more precise using incidence analysis.

First we use the formula for the impact of exogenous quantity on profits from Subsection 2.3.2. Let $\tilde{\rho}_n^{n+1}$ be the mark-up weighted version of $\hat{\rho}_n^{n+1}$. Then the impact of entry on profits of existing firms is bounded between the (inter-equilibria average of) the pre-entry profit incidence $(1 + \frac{n-1}{n}\rho)(P - c)q_{n+1}$ and the post-merger incidence $(1 + \frac{n}{n+1}\rho)(P - c)q_{n+1}$. The reduction in deadweight loss is bounded above by the pre-entry incidence formula, the average of $\frac{\rho}{n}(P - c)q_{n+1}$ and $\frac{\rho}{n+1}(P - c)q_{n+1}$.

¹³ They also establish weaker results when the integer constraints are included, but our results here are more precise than these.

The profit gained by entering is $(P - c)q_{n+1}$ evaluated at the post-merger equilibrium. Note that as we allow n to change continuously we recover the Mankiw and Whinston results as ρ , P and n are constant so the comparison of deadweight loss reduction to profit gain is just $\frac{\rho}{n}$ to 1. Discreteness makes entry more attractive to the extent that average mark-ups on the entry path are significantly above post-entry mark-ups, $\frac{n+1}{n}$ is significantly different from 1 and pass-through increases as prices fall (making its mark-up weighted average less than its straight weighted-average).

Thus incidence is a useful summary of the impact of entry on profits, quantities and social welfare. Because mergers in symmetric, constant returns Cournot industries are equivalent to exit, these could be analyzed in the same manner and combined naturally with marginal cost efficiencies, given that the impact of these will again be determined by incidence.

2.7.3 Price discrimination

A recent literature has revisited classical questions in the theory of monopoly price discrimination using an approach closely related to that employed here. Aguirre, Cowan, and Vickers (2010) (ACV) return to one of the oldest questions in industrial organization, posed by Pigou (1920): when does explicit third-degree price discrimination by a monopolist raise output and/or welfare?

Consider two markets, high (H) and low (L). Absent discrimination, prices are constrained to be identical. With discrimination, prices in H exceed those in L by Δ . ACV propose a natural continuous path from no discrimination to discrimination: we require that $p^H < p^L + \delta$. Assume profits in each market π^H and π^L are concave in price. Then for any $\delta \in [0, \Delta]$, the monopolist will choose $p^H = p^L + \delta$. Her first-order condition is thus $(\pi^H)'(p^L + \delta) + (\pi^L)'(p^L) = 0$. For $\delta < \Delta$, $(\pi^L)' < 0 < (\pi^H)'$, but

these both converge to 0 as δ goes to Δ .

A firm facing exogenous quantity \tilde{Q} earns profits $[Q(p) - \tilde{Q}](p - c)$. Her first-order condition is thus $Q'(p)(p - c) + Q(p) - \tilde{Q}$, while the first-order condition in the high market in the price discrimination problem is $(Q^H)'(p)(p - c) + Q^H(p) + (\pi^L)'(p - \delta)$. In effect, the downward pressure on prices from the constraint against discrimination in the low market enters in the same way as exogenous quantity. Moving towards discrimination is therefore equivalent to moving exogenous quantity from the high market to the low market.

Thus ACV show that discrimination leads to higher output if an average of (quantity) pass-through in the low market exceeds that in the high market. Similarly the change in social welfare in each market from the change in quantity is $\int M dq$ so a comparison of an average of the mark-up times the pass-through over the relevant range in the two markets determines the welfare effect of discrimination. The connections of pass-through to demand curvature make it clear how this result immediately implies the famous prior results of Pigou (1920), Robinson (1933), Schmalensee (1981) and Varian (1985) on the connections between demand curvature and the effects of discrimination.

Cowan (2010) analyzes the consumer surplus effects of discrimination by formulating the change from non-discrimination to discrimination instead in terms of a change in costs. Prior to discrimination, effectively costs in the low (high) market are elevated (depressed) by $\left| MR_n^{L(H)} - c \right|$ where $MR_n^{L(H)}$ represents the marginal revenue in the low (high) market at non-discriminatory prices. Introducing discrimination eliminates this cost shift and thus causes surplus in the low market to rise by $\bar{\rho} q_{nd}^L \Delta t$ and in the high market to fall by $\bar{\rho} q_{nd}^H \Delta t$, where this represents the average over the range between the discriminatory and non-discriminatory prices and Δt represents

the equivalent change in cost. Cowan argues that ρq is likely an increasing function of q (decreasing in p) so $\bar{\rho} q_{nd}^L > \rho_n^L q_n^L$, while $\bar{\rho} q_{nd}^H < \rho_n^H q_n^H$.

On the other hand at the non-discriminatory optimum, the marginal costs (excess of MC over MR in the high market) and benefits (excess of MR over MC in the low market) of lowering prices by an infinitesimal number of log points are equal and thus

$$\epsilon_n^L q_n^L (MR_n^L - c) = \epsilon_n^H q_n^H (c - MR_n^H),$$

where ϵ is the elasticity of demand. Thus, assuming Cowan's relations from the previous paragraph, a sufficient condition for discrimination to increase consumer welfare is that

$$\frac{\rho_n^L}{\rho_n^H} > \frac{\epsilon_n^L}{\epsilon_n^H}.$$

Cowan also derives necessary conditions using a similar approach.

The logic of incidence can be used to extend both of these results to the wide range of imperfectly competitive models we analyzed above. Consider the ACV result. Suppose that, with or without discrimination, each market is governed by the same conduct $l = \frac{p-c}{\mu}$ where μ is either independent or pooled depending on whether discrimination is allowed or not. Absent discrimination

$$\mu = -\frac{Q^H + Q^L}{(Q^H)' + (Q^L)'} = -\frac{Q^L - \frac{Q^H(Q^L)' - Q^L(Q^H)'}{(Q^H)' + (Q^L)'}}{(Q^L)'} = -\frac{Q^H + \frac{Q^H(Q^L)' - Q^L(Q^H)'}{(Q^H)' + (Q^L)'}}{(Q^H)'}$$

Thus the argument that the prohibition on discrimination acts as equal-and-offsetting exogenous quantity competition in the two markets in an amount $\frac{Q^H(Q^L)' - Q^L(Q^H)'}{(Q^H)' + (Q^L)'}$ holds generally. Because quantity pass-through in the two markets is $l\rho$, if l is constant in p , precisely the same results, interpreted in terms of averages of pass-through

or of demand curvature (as this is a simple transformation of pass-through), hold under imperfect competition. An analogous argument applies for Cowan's results.

Finally, Myatt and Rasmusen (2011) argue that the more infra-marginal surplus and deadweight loss there are, the more incentive a monopolist will have to perfectly price discriminate. To formalize this they assume that under third-degree price discrimination the monopolist must bilaterally Nash (1950) bargain with each buyer. Thus, if bargaining weights are equal, this is attractive if and only if the sum of consumer surplus and deadweight loss is greater than profits. As they observe, with constant returns, a sufficient condition is that demand be convex and a sufficient condition to avoid this is demand be concave, as these imply respectively pass-through globally above and below $\frac{1}{2}$. A generalization is that discrimination is desirable if $\bar{p} + \tilde{p} + 1 > \frac{1}{\lambda}$, where λ is the monopolist's bargaining weight.¹⁴

2.7.4 *Product design*

Spence (1975) considers a firm's optimal choice of a one-dimensional quality parameter, arguing that firms will move quality in a direction that redistributes from infra-marginal to marginal consumers. Johnson and Myatt (2006) study a special case of the Spence model where the quality parameter induces a rotation of the demand curve about a point and, under some restrictions of this parameterization, obtain sharper results. While illuminating, these exercises are all highly parametric in character. Here we show how the logic of incidence allows analysis of non-parametric formulations of these types of problems.

Suppose a monopolist must charge a uniform price and has normalized-to-zero

¹⁴ This generalization may or may not extend to case of general costs, depending on the interpretation one takes of bargaining with non-constant cost. It does not extend under the interpretation adopted by Myatt and Rasmusen.

variable cost of producing a good. She may face any demand such that potential gains from trade are constant, say at 1. What demand curves are most and least attractive to her?

Given that total potential gains from trade divide into consumer surplus, dead-weight loss and profits, the best curves will be those which minimize the ratio of the second two to the first and the worst those that maximize it. Thus her optimal demand curve has a pass-through rate everywhere as close to 0 as possible. This is achieved by the maximally concave demand curve, one which kinks from flat to vertical, namely there is a mass 1 of consumers each of whom value the good at 1. In this case the monopolist captures all surplus.

Her pessimal demand curve has pass-through rate every where as large as possible. For any fixed demand curve the monopolist can always make some profits by charging, say, the mean willingness to pay which is strictly positive and we can always construct another demand with higher pass-through than this one which will reduce her profits further. Thus there is no single pessimal demand curve. However, we can use Bulow and Pfleiderer (1983)'s class of constant pass-through demand curves to construct an increasing sequence that conveys the idea.

If demand is given by their form, $Q(p) = \sigma \left(1 + \frac{\rho-1}{\rho} \frac{p}{m}\right)^{\frac{\rho}{1-\rho}}$ then pass-through is constant at ρ and integration yields that consumer surplus is $\rho\sigma m \left(1 + \frac{\rho-1}{\rho} \frac{p}{m}\right)^{\frac{1}{1-\rho}}$. Thus when the price is efficient at 0 surplus is $\rho\sigma m$. Let m be constant at 1 and $\sigma = \frac{1}{\rho}$ for $\rho > 0$. We then obtain a parameterized class $\frac{\left(1 + \frac{\rho-1}{\rho} p\right)^{\frac{\rho}{1-\rho}}}{\rho}$. As $\rho \rightarrow \infty$ we have at each point a well-defined demand curve with total gains from trade of 1. However the monopolist's optimal price is 1 at every point and $Q(1) \rightarrow 0$. Thus her profits approach 0. Thus the sequence of arbitrarily bad demand curves approaches a constant elasticity demand curve, shifted vertically down one unit and scaled down

as the elasticity approaches 1 to preserve constant potential gains from trade. This is the maximally convex demand curve maintaining finite surplus and asymptotically denies all profit to the monopolist. Note that the particular sequence of demands we have chosen has also asymptotically 0 deadweight loss as the quantity below the monopoly price tends to 0.

While it would be excessive to here consider other problems of this sort in detail, we briefly mention two other possibilities and sketch their solution, highlighting the role of incidence:

1. What demand curve generates most waste? We want a curve that has low pass-through above the monopoly optimum but a high pass-through below. We could achieve this by mixing the two solutions above, having demand kink to being flat above the price of 1 but approach the shifted constant elasticity form at prices below 1.
2. What if the monopolist is taxed by a government (or supplied by an upstream firm) that seeks to maximize some combination of consumer surplus and the revenue it earns with no weight on the monopolist's profits? Now the monopolist's best demand curve from above is as bad as his worst as the authority can tax away all her profits. To see this, note that the government or upstream firm may charge the monopoly price to the monopoly without inducing it to raise price, thereby extracting all of its surplus. However, as we discuss further in the next subsection, high pass-through acts as a deterrent to taxation as it forces consumers to bear the burden of taxes, reducing tax revenue and increasing consumer surplus lost by taxation. Therefore the monopolist will want a high pass-through locally, but a low pass-through at higher prices (to avoid leaving too much surplus to consumers) and lower prices (to avoid too much dead-

weight loss), while still satisfying the second-order conditions of the upstream firm or authority.

The intuitions emerging from this analysis are in many ways similar to those of Spence and Johnson and Myatt and parametric versions of the variations on those problems. However, they show how incidence can provide a complementary non-parametric analysis of the incentives in product design that may be useful in settings where no particular parameterization is naturally compelling. Clearly, these exercises can be easily extended to imperfect competition.

2.7.5 *Supply chains*

A canonical model of supply chains proposed by Spengler (1950) has a “upstream” firm choosing its price, which is then taken by a downstream firm that charges prices to consumers. Natural extensions considered by many authors allow for multiple stages in the supply chain and imperfect competition at each stage.

First consider a supply chain consisting of several layers of imperfectly competitive firms supplying a necessary input to a downstream sector, which may then supply end-consumers or another downstream sector. There is a final demand function $Q(p_0)$ for the product. This, combined with a supply side structure, determines an equilibrium pass-through rate ρ_0 of the retailers as a function of the per-unit cost p_1 they are charged by the upstream firms. The upstream firms thus face effective demand $Q(p_0(p_1))$ with market power $\frac{\mu}{\rho_0}$, where $\mu \equiv -\frac{Q}{Q'}$ as before. Thus, letting l_1 be the competitiveness of the level 1 sector, equilibrium in an a symmetric upstream market is given by

$$p_1 - MC = \frac{\mu}{l_1 \rho_0}.$$

Thus the comparison of mark-ups between the upstream and downstream firms is given by the comparison of $l_1\rho$ to l_0 where l_0 is the competitiveness parameter downstream. This reasoning continues up the supply chain, with the aggregate pass-through of all levels beneath determining the incentives faced at each level. This implies that the pass-through from the n th to the $(n - 1)$ th level will depend on derivative of the pass-through from the $(n - 1)$ th to the n th level and thus on the $(2 + n)$ th derivative of demand, in principle allowing the identification from mark-up data of very high-order properties of demand, extending the logic of Villas-Boas and Hellerstein (2006). Conversely if constant pass-through is assumed many of these effects disappear, strong predictions are implied and the model is highly over-identified.

Analogous settings arise when firms sequentially choose how much of a homogeneous good to produce, as in the classic von Stackelberg (1934) model, extended by Anderson and Engers (1992) to the case when this occurs in many stages. The pass-through of quantities at each stage to the final market plays an analogous role to cost pass-through along a supply chain. Details are available on request.

2.7.6 *Aftermarkets and advertising*

Farrell (2008) considers a simple model of “aftermarkets” where a firm sells an add-on to a base product. Gabaix and Laibson (2006) and others have argued that often consumers fail to anticipate the expense of these add-ons. Farrell asks whether and by how much such failure harms consumers. This is very similar to the analysis of Dixit and Norman (1978) who study firms who, through advertising, can persuade consumers their product is worth more than it really is. However, Farrell is more concerned with whether a small amount of abusive after-product activity is harmful, while Dixit and Norman focus on whether at equilibrium too much or too little

advertising is supplied.

A small abusive after-market profit raises the effective profit of selling the initial good and thus acts as a uniform (if all consumers are homogeneous) subsidy on the product. Thus, to the first-order, the net effect of imposing a small such after-market “tax” on consumers that effectively subsidizes the initial good is $-Q(1 - \rho)$. Thus, as Farrell shows, such a subsidy may, paradoxically, actually increase consumer surplus if $\rho > 1$. The same holds for (a small bit of costless) advertising, except that the subsidy is given, effectively, to the consumers and thus the nominal price will rise by one unit more than in the Farrell case (though possibly still fall). Note, too, that social welfare clearly rises as long as $l > 0$ as total incidence is $Ql\rho$.

Now we consider the privately optimal level, in the spirit of Dixit and Norman.¹⁵ Suppose an industry association (or a monopolist) must choose a level of advertising a which uniformly increases willingness-to-pay for the product. Consumers lose $Q(1 - \rho)$ directly but also lose indirectly by purchasing a good they value less than they pay for it. This indirect loss is given, on the margin, by the Harberger triangle $-aQ'\rho$. Thus total consumer loss is $Q\left[1 - \rho\left(1 - \frac{a}{\mu}\right)\right]$. Thus, if firms in the industry perfectly collude on advertising levels, advertising is excessive (insufficient) if

$$1 > (<) \rho \left(1 - \frac{a}{\mu}\right).$$

Explicit assumptions about cost of advertising can tie down a further using the gains that firms make from advertising (from their incidence formula), but rather than continue on this analysis, we turn to an alternative perspective on advertising, that of Becker and Murphy (1993). In this case consumers gain the full benefit from the ad-

¹⁵ Our analysis differs from theirs, however, by allowing demand to be more general but advertising to enter demand in a more restricted way and using pre-advertising preferences to evaluate welfare.

vertisement and thus view advertisements as equivalent to quality improvements in the good or subsidies. Again sticking to the uniform benefit formulation, consumers gain $Q\rho$ from advertising, which is not internalized by the firm, and thus advertising is undersupplied at a rate proportional to ρ .

Of course, the uniform value increment assumption is a simplistic way to model the benefits from advertising. The distribution of marginal value increases from advertising could be modeled in many other ways, but incidence calculations for both the original demand and the marginal change from advertising would likely be crucial for the welfare properties in these cases as well.

2.7.7 *Taxing foreign trade*

Brander and Spencer (1981) ask when a government that does not care about the welfare of an imperfectly competitive sector will find it attractive to tax that sector to extract its rents.¹⁶ While Brander and Spencer motivate the problem with a focus on foreign trade, the same principles would apply to the regulation of a domestic firm if the planner does not value firm profits equally with consumer surplus and government revenues as in Laffont and Tirole (1993).

The government charges a specific tax t . If Q is the equilibrium demand of consumers, the marginal loss to consumers of the product is ρQ and to the government on infra-marginal tax is $t\rho Q'$, while the marginal revenue gain to the state is Q . Thus the optimum requires

$$1 = \rho \left(1 + \frac{t}{\mu} \right) \implies t^* = \mu \frac{1 - \rho}{\rho}.$$

¹⁶ They only consider a monopoly model, while we consider general imperfect competition.

Note that this formula in no way depends on the existence of imperfect competition; it applies equally well to the setting where the foreign firms are perfectly competitive. It thus unifies the analysis of Brander and Spencer with the classic analysis of terms-of-trade reasons for taxing imports as in Johnson (1953–1954). The only difference is that with an imperfectly competitive foreign sector, it is possible that $\rho > 1$ and thus a negative tax (subsidy) on imports may, in principle, be optimal. Thus the two theories are one, at least this far.

However, the externality of a tax on the foreign sellers is strictly greater with market power than in its absence: rather than $Q(1 - \rho)$, incidence on the foreign industry is $Q[1 - (1 - l)\rho]$. Thus there will be a stronger incentives for international trade agreements to limit such taxes between countries where firms have greater market power and in models where firms exercise this power, as in Ossa (2011).

2.7.8 Demand forms

As we hope has become clear from the examples above, incidence plays an important role in a range of problems in industrial economics. Parametric restrictions on it are thus powerful. The commonly-used linear, constant elasticity and exponential demand forms each imply constant pass-through of respectively $\frac{1}{2}$, $\frac{\epsilon}{\epsilon-1}$ and 1. In previous work (Weyl and Fabinger, 2009), we showed that most standard demand forms, even when flexible along many dimensions imply equally strong restrictions on pass-through and thus may not be flexible in the most relevant dimensions for many economic problems. We proposed a new *Adjustable pass-through* (Apt) demand function that allowed for flexibility on the level, elasticity, pass-through and slope of pass-through while maintaining most of the tractability of simpler demand forms in the Bulow and Pfleiderer (1983) constant pass-through class. In on-going work (Fabinger

and Weyl, 2012), we are using this form to obtain closed-form solutions canonical models, such as monopolistically competitive international trade models, and highlight the ways results obtained under particular incidence assumptions (arising from linear and/or constant elasticity demand models) depend on or are invariant to pass-through. We are also working on parameterized transformations of non-parametric demand functions that preserve, as with the stretch parameterization of Weyl and Tirole (2012), or transform the pass-through properties of an arbitrary demand function.

2.7.9 *An empirical example*

Atkin and Donaldson (2012) explicitly use the role of pass-through as a sufficient statistic and the structure of our results above to analyze the degree of competition in markets and the division of surplus from globalization. They consider markets for various internationally-traded commodities in different locations within developing countries in South Asia and Sub-Saharan Africa. They impose three key assumptions: that demand curvature is constant (demand is in the Bulow and Pfleiderer class) and the same across markets for a given product, returns are constant to scale and conduct is invariant to prices (l is constant). They then use the variation in empirical pass-through in the face of global price shocks across geographic locations for a given product to back out l , the degree of competition. Integrating and using the fact the under their assumptions local and global incidence are identical, they determine the division of surplus arising from the market existing between the intermediaries and consumers. This illustrates how relatively transparent restrictions on the behavior of pass-through can be used for empirical identification of questions of long-standing economic interest.

2.8 Conclusion

This paper argues that, just as in perfectly competitive models, incidence offers a powerful framework for organizing the analysis of comparative statics and welfare under imperfect competition. We have argued that, to paraphrase the conclusion of Bulow, Geanakoplos, and Klemperer (1985), the crucial question for welfare in imperfectly competitive markets is typically not “Do these markets exhibit price competition or quantity competition or competition using some other strategic variable?”, “Are products differentiated, how many firms are there, do firms act strategically or are they monopolistic competitors?” or even Bulow, Geanakoplos, and Klemperer’s “Do competitors think of the products as strategic substitutes or as strategic complements?” but rather, “What is the pass-through and incidence of a tax in this market?” Unlike the first group of questions, including the Bulow, Geanakoplos, and Klemperer formulation this last question is not “new” to oligopoly theory. Rather it is what, at least since the time of Marshall, economists have been asking about competitive markets to analyze a wide range of outcomes and policies. Thus the analysis of “strategic” industries with market power may not be as distinct as it may at first seem from the analysis of perfectly competitive markets.

While we offered a number of applications of incidence reasoning we neither exhausted potential applications nor exhaustively treated those we considered. More basic theoretical work remains to be done on incidence, particularly in imperfectly competitive asymmetric differentiated products industries. We briefly discuss each of these directions for future in this conclusion.

Extensions of many of the applications to related models could be fruitful, such as:

- Other models of platforms with unidimensional heterogeneity, such as those of Anderson and Coate (2005), Armstrong (2006), Goos, Cayseele, and Willekens (2011) and Gomes and Pavan (2011).
- Enrichments of the procurement models of Section 2.6 to combine aspects of competition, moral hazard, product selection and ex-post monitoring.
- Merger and entry analysis in non-symmetric, differentiated products industries with more than two firms.

Other promising applications of incidence reasoning include:

- The design and analysis of auctions, some connections to which are drawn by Mares and Swinkels (2011).
- Nonlinear pricing and optimal taxation, where curvature of distributions was shown by Saez (2001) to play an important role.
- Almost all international trade models use explicit, often constant pass-through demand forms to obtain results, which are known to vary based on, for example, whether linear or constant elasticity demand are employed. It thus seems likely that incidence plays an important role in the comparative statics of such models.

Finally, there are some important areas for extending the basic logic of incidence. Generalizations to multi-product firms and to relax the assumptions of concavity and stability made here were included in previous versions of this paper and are available on request. But the analysis of asymmetric, differentiated and imperfectly competing firms, a workhorse of empirical industrial organization, remains under analyzed; see

Gabaix, Laibson, Li, Li, Resnick, and de Vries (2010), Quint (2012), earlier versions of this paper and Jaffe and Weyl (2012) for some suggestive but special results about the role of incidence in these settings.

3. PRICE DYNAMICS FOR DURABLE GOODS

3.1 *Introduction*

Expenditure on durable goods accounts for 60% of consumption expenditure and all of investment expenditure. It is the most volatile component of GDP at business cycle frequencies. A large fraction of international trade is in durable goods. As is well known, the durable nature of the product makes the pricing of these goods differ from that of nondurables, since consumers demand for durables today depends not only on prices today but also on their expectation of future prices. In the vast majority of macroeconomic models (both closed and open-economy) that study the behavior of durables pricing of durables is treated similar to that of nondurables: either perfect competition is assumed and firms are price takers or if the firm has pricing power they are assumed to not internalize the effect of their price today on future demand for their product, despite the durable nature of their product (Barsky, House, and Kimball (2007)). In the open economy literature the impact of exchange rate movements on prices is studied in environments where goods are assumed to be nondurable, despite the preponderance of durable goods in trade.

On the other hand there is a large microeconomic literature that studies the specific problem of pricing of durable goods and its special features relative to non-durable pricing. The forward looking nature of demand implies that the prices firms set will depend on whether they can commit to future prices or cannot. In one of the

earliest papers in the literature Coase (1972) conjectured that in the absence of commitment a monopolistic firm producing durable goods will be bound to charge the marginal cost due to perfect inter-temporal competition with itself. This conjecture has been proven by Stokey (1981), Bulow (1982) and Gul, Sonnenschein, and Wilson (1986) in various setups, as discussed in the Supplementary Section 1.5 of Chapter 1 of Tirole (1988). This conjecture is a limiting result in an environment where prices adjust at each instant (continuous time) or there is zero depreciation of the durable good. A large literature has followed as surveyed in Waldman (2003). Much of this analysis has focused on long-run pricing behavior with less emphasis on dynamics and response to shocks. Also, the analysis is typically done for the case of a monopolist. As Waldman (2003) mentions in his conclusion (see p. 150) “most of the literature assumes either monopoly or perfect competition, while clearly most real world markets are either oligopolistic or monopolistically competitive”.

In this work we bring the insights in the microeconomic literature on durable goods pricing into macroeconomic environments. Consistent with macroeconomic treatment of durable goods we allow for positive depreciation rates and discrete time periods between price setting and evaluate several pricing environments including monopoly, oligopoly and monopolistic competition. We explore the cases of commitment and discretion and evaluate the response to cost and demand shocks.

We consider a partial equilibrium environment and focus on firms price setting given a consumer demand function that depends on its prices today and in the future. In the case with commitment, prices are independent of the past levels of durable goods consumption and of past prices. It depends on the current level of demand and on current and future costs. If the elasticity of demand (with respect to the “rental” price of the durable good) is constant then prices are a constant markup over

marginal costs. There are no endogenous dynamics in prices. There is complete and instantaneous pass-through of cost (exchange rate shocks) into prices and demand shocks have no effect on prices. When firms have the ability to commit they are able to commit not to compete with themselves and thus obtain monopoly rents.

As is well known the commitment solution is not dynamically consistent. The demand for the durable good depends on its expected future price. In the current period the monopolist would benefit if the consumers believed that the future prices of the durable good would be high, but in the next period, the monopolist would like to lower the price in order to increase sales. In the absence of the monopolist's ability to commit to high future prices, the consumers will base their current purchases on their expectation of low future prices. This impedes the producer's ability to capture the full potential monopoly rent.

We evaluate the implications of the time consistent solution for the dynamics of adjustment to cost and demand shocks and focus on Markov perfect equilibria. In this environment prices depends on the endogenous state variable, the stock of durables. Consequently prices adjust sluggishly to shocks even when the shock is a permanent shock. Markups move endogenously over time. In response to a positive cost shock (exchange rate shock) prices increase but by less than the percent increase in costs and markups decline generating incomplete pass-through. When costs increase firms mute the price response to prevent consumers from shifting demand to the future when they cannot commit to keep prices high (relative to their marginal cost). Demand shocks also now impact pricing. Markups and prices increase in response to positive demand shocks.

The fact that markups decrease in response to a positive cost shock has implications for the literature on incomplete exchange rate pass-through. Most traded goods

are durable in nature. The fact that pass-through is incomplete in the long-run is frequently attributed to strategic complementarities in pricing that prevents a firm from raising its price in response to cost shocks as this causes the elasticity of demand it faces to rise. Adding the assumption of frictions in price adjustment then generates dynamics in pass-through.

Therefore, in the case of durable goods, with discretion in pricing one obtains pass-through dynamics even in the flexible price case, and incomplete pass-through even in the absence of standard strategic complementarities in pricing and constant elasticity of demand. This contrasts with the literature on exchange rate pass-through that treats all goods as nondurable and where endogenous dynamics in prices arise because of infrequent price adjustments and pass-through is incomplete in the long-run because of variable markups that arise from strategic complementarities in pricing. The solution here also contrasts with the standard macroeconomic assumption of marginal cost pricing of durable goods. The endogeneity of markups implies that, in response to cost shocks, the volatility in prices and therefore quantity is lower than the case of constant markup pricing.

In the case of oligopolistic competition where firms engage in Cournot competition in producing a homogenous good the pricing decision of each firm is influenced by two different forces: competition with the other firms and inter-temporal competition with itself. We show here that the extent of dynamics in prices, in the time-consistent solution, is a decreasing function of the number of firms. That is, with a large number of competing firms, the across-firm competition dominates the inter-temporal within firm competition of the firm. In the case of monopolistic competition where each firm produces an individual variety and is assumed to be infinitesimally small relative to the industry we show that the problem is similar that of a durable

good monopolist subject to stochastic demand shocks. We also show that the price of a particular variety (and the persistence in price dynamics) of durable good depends more strongly on the deviation of the average stock of durables in the industry from its steady state as compared to the deviations of the stock of its variety relative to the industry average.

In future versions of this work we plan to nest the durable goods pricing problem into a general equilibrium macroeconomic environment. A paper in the literature that speaks to the dynamic response to demand shocks is Caplin and Leahy (2006). They provide an (S, s) model of durable stock adjustment by heterogeneous consumers with monopoly pricing by firms, also in a partial equilibrium environment. Our approach to modeling demand is very different which allows us to address the cases of oligopoly and monopolistic competition. Esteban and Shum (2007) study price and quantity dynamics in an oligopolistic environment with secondary markets for the case of automobiles. Their focus is on measuring the competitive importance of the secondary market.

Section 3.2 describes the demand for durable goods. Section 3.3 derives results for pricing and quantity dynamics for the case of a monopolist and Sections 3.4 and 3.5 analyze dynamics for the case of oligopoly and monopolistic competition.

3.2 *Durable good demand*

Consider an infinitesimal agent deriving instantaneous utility $U(C_t, D_t; \xi_t)$ from consumption of durable good D_t and nondurable good C_t in period t . The parameter ξ_t represents a demand shock. The time is discrete and the one-period discount factor is β . Period t purchases of the durable good are denoted by X_t and δ is the depreciation rate of the durable, so the dynamics of the stock of durable is described

by

$$D_t = (1 - \delta)D_{t-1} + X_t. \quad (3.1)$$

Denote by P_t the price of the durable good and by P_{Ct} the price of the nondurable good, which is being consumed in positive quantities each period. The agent faces one-period gross interest rate equal to $1/\beta$. The agent maximizes

$$E_0 \sum_{t=0}^{\infty} \beta^t U(C_t, D_t; \xi_t)$$

subject to

$$\sum_{t=0}^{\infty} \beta^t (P_{Ct}C_t + P_t (D_t - (1 - \delta)D_{t-1})) \leq NPV_0$$

where NPV_0 represents the discounted net present value of the agent's resources¹.

The first order conditions for the choice of C_t and D_t are

$$U_1(C_t, D_t; \xi_t) = \lambda_t P_{Ct}$$

and

$$U_2(C_t, D_t; \xi_t) - \lambda_t P_t + \beta(1 - \delta)\mathbb{E}_t \lambda_{t+1} P_{t+1} = 0$$

where λ_t is the Lagrange multiplier on the period t budget constraint. In deriving the first order condition we have implicitly assumed that there are no irreversibility of purchases constraints. In other words, the consumer can always sell the remaining stock of the durable good at the market price. Alternatively, we may assume that shocks are small and prices are never so high that the irreversibility constraint $X_{t+1} \geq 0$ binds. That is, the representative consumer will want to purchase positive amounts

¹ Generalization to stochastic consumer's endowment is straightforward.

of the durable good in every period. This will be the case, for example, if δ is high enough.

Let us now assume that the period utility function is quasi-linear, $U(C_t, D_t; \xi_t) \equiv C_t + u(D_t; \xi_t)$, and normalize $P_{Ct} = 1$. The first order conditions are then equivalent to $\lambda_t = 1$, and²

$$u'(D_t; \xi_t) = P_t - \beta(1 - \delta)\mathbb{E}_t P_{t+1} \quad (3.2)$$

The marginal utility from an additional unit of the durable good should equal the price of the durable net of the expected future resale value of the undepreciated durable goods stock. In the following sections we use this demand equation as a starting point, specializing at times to *linear* or *constant elasticity* demand. Linear demand corresponds to a quadratic utility function u . With a convenient choice of the demand shock ξ_t , the marginal utility may be written as

$$u'(D_t; \xi_t) = a + \xi_t - bD_t.$$

for some parameters a and $b > 0$. Constant elasticity demand³ corresponds to u being a concave power function. The demand shock ξ_t is chosen so that

$$u'(D_t; \xi) = \xi_t D_t^{-1/\sigma},$$

where σ is the demand elasticity in case of no durability, i.e., $\delta = 1$.

In the next section we introduce the firm that produces the durable good and its

² In general, for functions of multiple variables, prime is used to denote the partial derivative with respect to the first argument.

³ This choice of terminology requires caution. The demand function does not have constant elasticity with respect to the price P_t , but with respect to the price at which the durable goods could be, in principle, rented, namely $P_t - \beta(1 - \delta)\mathbb{E}_t P_{t+1}$.

price setting decision. We assume that the firm's production function is linear, with constant marginal costs W_t that can vary over time.

3.3 Durable good monopoly

3.3.1 The commitment case as a benchmark

Consider a monopolistic firm producing durable goods at a marginal cost W_t in period t , which can commit to a sequence of prices $\{P_t\}_{t=0}^{\infty}$ of its choice. The prices may depend on the state of the world. The sequence of prices is chosen in period $t = 0$ to maximize the discounted net present value of profits

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t (P_t - W_t) X_t,$$

where X_t satisfies (3.1) and D_t satisfies (3.2). For convenience, we denote by $(1 - \delta)D_{-1}$ the stock of the durable good the consumers had at the beginning of period $t = 0$ before making any purchases in that period. The following lemma characterizes the optimal choice of prices by the monopolist.

Lemma 1 (Monopoly, pricing with commitment): Provided that the initial condition is $D_{-1} = 0$, the durable good monopolist's optimal pricing with commitment satisfies

$$P_t = \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j (1 - \delta)^j \frac{\sigma_{t+j}}{\sigma_{t+j} - 1} (W_{t+j} - \beta(1 - \delta)\mathbb{E}_{t+j}W_{t+j+1}),$$

where

$$\sigma_t \equiv \sigma(D_t; \xi_t) = -\frac{u'(D_t; \xi_t)}{u''(D_t; \xi_t)D_t},$$

and the level of demand D_t is determined by

$$u'(D_t; \xi_t) = \frac{\sigma_t}{\sigma_t - 1} (W_t - \beta(1 - \delta)\mathbb{E}_t W_{t+1})$$

Proof: The Lagrangian for the firm's problem may be written as

$$\mathcal{L}_0 = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t [(P_t - W_t) (D_t - (1 - \delta)D_{t-1}) + \lambda_t (u'(D_t; \xi_t) - P_t + \beta(1 - \delta)P_{t+1})],$$

where λ_t is the Lagrange multiplier for the demand condition (3.2). The optimality conditions are

$$\begin{aligned} (D_t - \lambda_t) - (1 - \delta)D_{t-1} &= 0 \quad \text{for } t = 0, \\ (D_t - \lambda_t) - (1 - \delta)(D_{t-1} - \lambda_{t-1}) &= 0 \quad \text{for } t > 0, \\ (P_t - W_t) - \beta(1 - \delta)\mathbb{E}_t (P_{t+1} - W_{t+1}) + \lambda_t u''(D_t) &= 0 \quad \text{for all } t. \end{aligned}$$

When $D_{-1} = 0$, the optimal path of D_t is $D_t = \lambda_t$, and the optimal path of P_t satisfies

$$(P_t - W_t) - \beta(1 - \delta)\mathbb{E}_t (P_{t+1} - W_{t+1}) = -D_t u''(D_t) \quad (3.3)$$

Using the demand equation (3.2) and rearranging terms yields the demand condition in the proposition:

$$u'(D_t; \xi_t) = \frac{\sigma_t}{\sigma_t - 1} (W_t - \beta(1 - \delta)\mathbb{E}_t W_{t+1})$$

Recursively substituting for P_{t+j} in (3.2), letting $\mathbb{E}_t \beta^j (1 - \delta)^j P_{t+j} \rightarrow 0$ as $j \rightarrow \infty$, and substituting for $u'(D_t; \xi_t)$ using the previous equation one arrives at the price equation in the proposition. ■

Note that in the case of competitive pricing, $P_t = W_t$, and D_t evolves according to $u'(D_t; \xi_t) = W_t - \beta(1 - \delta)\mathbb{E}_t W_{t+1}$, which differs from the monopoly pricing solution by the factor $\sigma_t / (\sigma_t - 1)$.

In general, it follows from the proposition that there is no endogenous dynamics in prices. Prices set at time t do not depend on lagged prices, only on current and future costs. If the costs W_t follow an AR(1) process with mean \bar{W} and persistence ϕ_W , the price sequence satisfies

$$P_t = (1 - \beta(1 - \delta)) \sum_{j=0}^{\infty} \beta^j (1 - \delta)^j \mathbb{E}_t \left\{ \frac{\sigma_{t+j}}{\sigma_{t+j} - 1} \left[\bar{W} + (W_{t+j} - \bar{W}) \frac{1 - \beta(1 - \delta)\phi_W}{1 - \beta(1 - \delta)} \right] \right\}.$$

For constant elasticity demand, $\sigma_t \equiv \sigma$, and arbitrary processes for cost and demand shocks, the optimal pricing policy is the standard constant markup pricing:

$$P_t = \frac{\sigma}{\sigma - 1} W_t.$$

This pricing has the following properties: (a) complete instantaneous pass-through; (b) independence from demand shocks; and (c) independence from the depreciation rate δ and length of time period (inversely related to β). It coincides with the optimal nondurable pricing policy. This will be our benchmark for comparison in what follows.

With linear demand $u'(D; \xi) = a + \xi - bD$ and AR(1) processes for demand and cost shocks (with persistence parameters ϕ_ξ and ϕ_W), we have

$$\begin{aligned} a + \xi_t - 2bd_t &= \bar{W} + (W_t - \bar{W})[1 - \phi_W\beta(1 - \delta)], \\ \sigma_t &= \frac{a + \xi_t}{bD_t} - 1, \\ P_t &= \frac{1}{2} \left[\frac{a}{1 - \beta(1 - \delta)} + (W_t - \bar{W}) + \frac{\xi_t}{1 - \phi_\xi\beta(1 - \delta)} \right]. \end{aligned}$$

Here the mean of ζ_t is set to zero.

3.3.2 *Discretionary pricing of durable goods*

It is well known that the commitment solution for the monopolist is time inconsistent. The demand for the durable good depends on its expected future price. In the current period the monopolist would benefit if the consumers believed that the future prices of the durable good would be high, but in the next period, the monopolist would like to lower the price in order to increase sales. In the absence of the monopolist's ability to commit to high future prices, the consumers will base their current purchases on their expectation of low future prices. This impedes the producer's ability to capture the full potential monopoly rent.

In this subsection we evaluate the case of such discretionary pricing. In each period the monopolist sets prices taking as given the residual demand for the good. The monopolistic firm does not internalize the effect its expected pricing in the current period had on demand in the previous period. We solve for Markov perfect equilibria.

In the case under consideration, the demand for the product in each period is still given by (3.2), where $\mathbb{E}_t P_{t+1}$ are the expectations of consumers about the pricing policy of the monopolist. There are two exogenous state variables W_t and ζ_t , both of which follow Markov processes. We restrict attention to Markov perfect equilibria where the only endogenous state variable is D_{t-1} . In other words, we assume that only the current level of the 'physical' state variable matters for decisions economic agents, not the full history leading to this level. A change in D_t affects the policy of the firm in future periods and consumers take this into account when they demand the durable good today.

Formally, denote by $V(D_{t-1}; W_t, \xi_t)$ the value of the firm as a function of the endogenous and exogenous state variables. The value satisfies the Bellman equation

$$V(D_{t-1}; W_t, \xi_t) = \max_{P_t, D_t} \{ (P_t - W_t) (D_t - (1 - \delta)D_{t-1}) + \beta \mathbb{E}_t V(D_t; W_{t+1}, \xi_{t+1}) \} \quad (3.4)$$

subject to the demand condition

$$u'(D_t; \xi_t) = P_t - \beta(1 - \delta)\mathbb{E}_t P_{t+1}, \quad (3.5)$$

We represent the optimal behavior of the firm and the consumers using policy functions d , p , and f as follows:

$$\begin{aligned} P_t &= p(D_{t-1}; W_t, \xi_t) \\ D_t &= d(P_t; W_t, \xi_t) = d(p(D_{t-1}; W_t, \xi_t); W_t, \xi_t) \equiv f(D_{t-1}; W_t, \xi_t) \end{aligned}$$

The first order condition for the choice of P gives

$$\begin{aligned} D_t - (1 - \delta)D_{t-1} + (P_t - W_t) d'(P_t; W_t, \xi_t) \\ + \beta d'(P_t; W_t, \xi_t) \mathbb{E}_t V'(d(P_t; W_t, \xi_t); W_{t+1}, \xi_{t+1}) = 0, \end{aligned}$$

which combined with the envelope condition

$$V'(D_{t-1}; W_t, \xi_t) = -(1 - \delta) (P_t - W_t)$$

yields

$$(D_t - (1 - \delta)D_{t-1}) + [(P_t - W_t) - \beta(1 - \delta)\mathbb{E}_t (P_{t+1} - W_{t+1})] d'(P_t; W_t, \xi_t) = 0. \quad (3.6)$$

Substituting $d(P_t; W_t, \xi_t)$ for D_t and $p(d(P_t; W_t, \xi_t); W_{t+1}, \xi_{t+1})$ for P_{t+1} in (3.5) and differentiating with respect to P_t leads to

$$d'(P_t; \xi_t) = \frac{1}{u''(D_t; \xi_t) + \beta(1 - \delta)\mathbb{E}_t p'(D_t; W_{t+1}, \xi_{t+1})} \quad (3.7)$$

The following lemma characterizes the equilibrium pricing policy without commitment.

Lemma 2 (Monopoly, pricing without commitment): (i) The dynamics of prices and quantities of the durable good are described by the following dynamic system:

$$(P_t - W_t) - \beta(1 - \delta)\mathbb{E}_t (P_{t+1} - W_{t+1}) \quad (3.8)$$

$$= (D_t - (1 - \delta)D_{t-1}) (u''(D_t; \xi_t) + \beta(1 - \delta)\mathbb{E}_t p'(D_t; W_{t+1}, \xi_{t+1})),$$

$$u'(D_t; \xi_t) = p(D_{t-1}; W_t, \xi_t) - \beta(1 - \delta)\mathbb{E}_t p(D_t; W_{t+1}, \xi_{t+1}). \quad (3.9)$$

This system simultaneously determines the equilibrium dynamics of D_t and the optimal policy function p .

(ii) The policy functions satisfy $p'(D_{t-1}; W_t, \xi_t) < 0$ and $f'(D_{t-1}; W_t, \xi_t) > 0$.

Proof: Equations (3.8) and (3.9) follow from (3.6), (3.7), and (3.5). The conditions $p'(D_{t-1}; W_t, \xi_t) < 0$ and $f'(D_{t-1}; W_t, \xi_t) > 0$ are derived in the appendix. ■

Condition (3.8), or (3.6), has an intuitive interpretation: it is the firm's optimality condition that can be obtained using a perturbation argument. We can rewrite it as

$$P_t - \beta(1 - \delta)\mathbb{E}_t P_{t+1} + \frac{1}{d'(P_t; W_t, \xi_t)} X_t = W_t - \beta(1 - \delta)\mathbb{E}_t W_{t+1}.$$

The left hand side is the marginal revenue associated with increasing the quantity

$X_t = D_t - (1 - \delta)D_{t-1}$ sold in period t and reducing the quantity X_{t+1} sold in the next period in a way that leaves D_{t+1} unchanged. The first two terms represent the direct gain from markups, which would be present even in competitive markets. The third term captures the loss in the monopolist's profit margin due to the negative movement in price P_t . The right hand side, of course, represents the marginal cost corresponding to this small change in production.

Lemma 2 fully characterizes the equilibrium dynamics. The equilibrium pricing is suboptimal in the sense that it no longer satisfies the conditions of Lemma 1. To see this observe the two extra expressions in (3.8) compared to (3.3): D_{t-1} and $p'(D_t; W_{t+1}, \xi_{t+1})$. Both are in general non-zero and they are directly related to the time inconsistency of the monopolistic pricing. In particular, D_{t-1} appears because the monopolist does not internalize the effect of the pricing policy on demand last period. The expression $p'(D_t; W_{t+1}, \xi_{t+1})$ represents the fact that the monopolist can partially affect its future pricing policy through the state variable.

Result 3 *In the absence of shocks to cost and demand ($W_t = \bar{W}$, $\xi_t = \bar{\xi}$) the steady state price is given by*

$$\bar{P} = \frac{\bar{\sigma}}{\bar{\sigma} - \delta\bar{\omega}} \bar{W}$$

where $\bar{\sigma} \equiv \sigma(\bar{D}; \bar{\xi})$, steady state consumption \bar{D} satisfies $u'(\bar{D}) = (1 - \beta(1 - \delta)) \bar{P}$, and $\bar{\omega} \equiv 1 + \beta(1 - \delta)p'(\bar{D}; \bar{W}, \bar{\xi})/u''(\bar{D}; \bar{\xi})$. As a corollary, there is marginal cost pricing in the long run if the good is perfectly durable, i.e., $\bar{P} = \bar{W}$ when $\delta = 0$.

Proof: This result directly follows from (3.8)-(3.9). ■

Note that for $\delta > 0$ the value of steady state markup is related to the out-of-steady-state behavior of prices.

Solution in the case of quadratic utility

Bond and Samuelson (1984) studied the durable good monopolist problem in the case of quadratic utility. They demonstrated the following properties of the steady state: (i) in the case of zero depreciation rate, marginal cost pricing is reached in the long run; (ii) for a non-zero depreciation rate, price is above marginal cost in the long-run if the time period has non-zero length; (iii) in the limit of zero length of the time period, the durable good price is equal to the marginal cost, independently of δ , i.e., the Coase conjecture holds. In the present work, we additionally study the dynamics of pricing.

Lemma 4 (Linear pricing under quadratic utility): Let W_t and ξ_t follow AR(1) processes with means \bar{W} and zero, respectively. With quadratic utility, $u'(D_t; \xi_t) = a - bD_t + \xi_t$, there exists a linear solution to the pricing equations (3.8)-(3.9):

$$p(D_{t-1}; W_t, \xi_t) = \bar{P} - \alpha(D_{t-1} - \bar{D}) + \gamma(W_t - \bar{W}) + \lambda\xi_t, \quad (3.10)$$

In addition, the dynamics of the state variable is also linear and satisfies

$$D_t = \bar{D} + \phi(D_{t-1} - \bar{D}) - \psi(W_t - \bar{W}) + \eta\xi_t. \quad (3.11)$$

Here $\alpha, \gamma, \lambda, \phi, \psi$, and η are constants.

Proof: In this case $u''(D_t; \xi_t) = -b$. Substituting the conjectured linear policy rules (3.10) and (3.11) into the pricing equations (3.8) and (3.9) and comparing coefficients of different terms eight conditions for the eight unknown parameters. The solution

to this system of equations is

$$\begin{aligned}
\alpha &= \frac{b(\zeta-1)}{\beta(1-\delta)} > 0, & \phi &= \frac{\zeta-1}{\zeta\beta(1-\delta)} \in (0,1), \\
\gamma &= \frac{\zeta}{1+\zeta} \in (1/2,1), & \psi &= \frac{1-\beta(1-\delta)\phi_W}{b(1+\zeta)} > 0, \\
\lambda &= \frac{1}{1+\zeta} \frac{1}{1-\beta(1-\delta)\phi_\zeta} > 0, & \eta &= \frac{1}{(1+\zeta)b} > 0, \\
\bar{P} &= \frac{W}{1-\beta(1-\delta)+\delta\zeta} + \frac{a}{1-\beta(1-\delta)} \frac{-\beta(1-\delta)+\delta\zeta}{1-\beta(1-\delta)+\delta\zeta}, & \bar{D} &= \frac{1}{b} \frac{a-(1-\beta(1-\delta))W}{1-\beta(1-\delta)+\delta\zeta}.
\end{aligned}$$

where $\zeta \equiv (1 - \beta(1 - \delta)^2)^{-1/2} > 1$, and ϕ_W and ϕ_ζ are the persistence parameters of the process W_t and ζ_t , respectively. It is straightforward to verify that with these parameters, the dynamic equations are satisfied. ■

Note that equations (3.10) and (3.11) imply the following process for prices:

$$\begin{aligned}
P_t - \bar{P} &= \phi(P_{t-1} - \bar{P}) + \gamma(W_t - \bar{W}) + \lambda\zeta_t - \phi\beta(1-\delta)\gamma\phi_W(W_{t-1} - \bar{W}) \\
&- \phi(1 + \lambda\beta(1-\delta)\phi_\zeta)\zeta_{t-1}.
\end{aligned}$$

Result 5 *In this environment prices adjust slowly over time in response to shocks (even one-time permanent shocks) and pass-through is incomplete. As the durability of the good declines ($\delta \rightarrow 1$), the persistence parameter ϕ approaches zero, i.e., in the nondurable limit there is no endogenous dynamics.*

1. D_t increases over time, while P_t and markup decrease over time.
2. Markup and price increase in response to a positive demand shock (procyclical markup).
3. Markup decreases and price increases in response to a positive cost shock (countercyclical markup).

The fact that markups decrease in response to a positive cost shock has implications for the literature on incomplete exchange rate pass-through. Most traded goods

are durable in nature. The fact that pass-through is incomplete in the long-run is frequently attributed to strategic complementarities in pricing that prevents a firm from raising its price in response to cost shocks as this causes the elasticity of demand it faces to rise. Adding the assumption of frictions in price adjustment then generates dynamics in pass-through.

We have seen that in the case of durable goods with discretion in pricing one obtains pass-through dynamics even in the flexible price case, and incomplete pass-through even in the case the absence of standard strategic complementarities in pricing. In a sense, in the case under consideration, there are strategic complementarities over time that arise from the firm competing with itself.

The solution here also contrasts with the marginal cost pricing case of durable goods. It exhibits smaller price volatility and quantity volatility.

General utility functions

Here we present a compact equation that determines the transition function f in the absence of demand shocks and cost shocks. Note that for small deviations from the steady state \bar{D} the persistence parameter is given by $f'(\bar{D})$.

Lemma 6 (Transition function for general utility): The transition function f , for any D , satisfies

$$\begin{aligned} & \frac{(1 - \beta(1 - \delta)) W - u'(f(D))}{f(D) - (1 - \delta) D} - u''(f(D)) \\ &= \beta(1 - \delta) \frac{(1 - \beta(1 - \delta)) W - u'(f(f(D)))}{f(f(D)) - (1 - \delta) f(D)} f'(f(D)). \end{aligned} \tag{3.12}$$

Moreover, with the knowledge of f , the consumer reaction function d may be recov-

ered as the solution to

$$f(d(P)) = d\left(\frac{P - u'(d(P))}{\beta(1-\delta)}\right) \quad (3.13)$$

Proof: Notice that (3.13) is an immediate consequence of the demand equation $P - u'(d(P)) = \beta(1-\delta)p(d(P))$. We just need to demonstrate the result (3.12). We start with (3.6) adapted to the case under consideration, with time indices suppressed:

$$\left(d(P) - (1-\delta)f^{-1}(d(P))\right) + [(P - W) - \beta(1-\delta)(p(d(P)) - W)]d'(P) = 0.$$

The demand equation $P - \beta(1-\delta)p(d(P)) = u'(d(P))$ then implies

$$d(P) - (1-\delta)f^{-1}(d(P)) = [(1-\beta(1-\delta))W - u'(d(P))]d'(P).$$

Applying the inverse function differentiation theorem $d^{-1'}(D) = 1/d'(d^{-1}(D))$, we get

$$d^{-1'}(D) = \frac{(1-\beta(1-\delta))W - u'(D)}{D - (1-\delta)f^{-1}(D)}.$$

Integrating both sides of this equation gives

$$d^{-1}(D) - \bar{P} = \int_{\bar{D}}^D \frac{(1-\beta(1-\delta))W - u'(\tilde{D})}{\tilde{D} - (1-\delta)f^{-1}(\tilde{D})} d\tilde{D}. \quad (3.14)$$

This equation must hold for any D . Using this last equation at a different point, and noting that $d^{-1}(D_+) = p(D)$, we get

$$p(D) - \bar{P} = \int_{\bar{D}}^{D_+} \frac{(1-\beta(1-\delta))W - u'(\tilde{D})}{\tilde{D} - (1-\delta)f^{-1}(\tilde{D})} d\tilde{D}. \quad (3.15)$$

Multiplying (3.15) by $-\beta(1-\delta)$ and adding the resulting equation to (3.14) yields

$$u'(D) - (1 - \beta(1 - \delta))\bar{P} \\ = \int_{\bar{D}}^D \frac{(1 - \beta(1 - \delta))W - u'(\tilde{D})}{\tilde{D} - (1 - \delta)f^{-1}(\tilde{D})} d\tilde{D} - \beta(1 - \delta) \int_{\bar{D}}^{f(D)} \frac{(1 - \beta(1 - \delta))W - u'(\tilde{D})}{\tilde{D} - (1 - \delta)f^{-1}(\tilde{D})} d\tilde{D}.$$

Here we simplified the left hand using the demand equation $d^{-1}(D) - \beta(1 - \delta)p(D) = u'(D)$. Differentiation with respect to D (and shifting the time period under consideration) leads to the final result (3.12). ■

Numerical solution

We further study the properties of the dynamics of durable good pricing without commitment in the special case of constant elasticity utility by means of a numerical solution. The solution to the dynamic system (3.8)–(3.9) does not have a simple characterization in this case, and we obtain the model's solution using value function iteration.

Specifically, we start with a guess for the value function, $V(\cdot)$, and pricing rule, $p(\cdot)$. Given the pricing rule, we can use demand (3.2) express current period price as a function of current period stock D :

$$P = u'(D) + \beta(1 - \delta)\mathbb{E}p(D),$$

where we have suppressed demand and cost shocks from notation (including in the expectation). Substituting this expression into the value function, we can write the

value of the firm recursively:

$$\tilde{V}(D_-) = \max_D \left\{ (u'(D) + \beta(1 - \delta)\mathbb{E}p(D) - W)(D - (1 - \delta)D_-) + \beta\mathbb{E}V(D) \right\}, \quad (3.16)$$

where D_- is the state variable—the stock of the durable good last period. The arg max of this problem is the state variable transition function, $D = \tilde{f}(D_-)$. From this we can update the pricing rule according to:

$$\tilde{p}(D_-) = u'(\tilde{f}(D_-)) + \beta(1 - \delta)\mathbb{E}p(\tilde{f}(D_-)).$$

Hence, on each iteration, given the initial guesses $V(\cdot)$ and $p(\cdot)$, we obtain a new value function $\tilde{V}(\cdot)$ and a new pricing rule $\tilde{p}(\cdot)$. We repeat this procedure until convergence. In order to obtain convergence, we apply polynomial smoothing to $\tilde{V}(\cdot)$ and $\tilde{p}(\cdot)$ on each iteration.

In this numerical solution, we focus on the case of constant elasticity demand with $\sigma = 2$. Larger values of σ mute the endogenous dynamics of durable monopolist prices. We choose the remaining two parameters—the discount rate and depreciation rate—at $\beta = 0.9$ and $\delta = 0.2$ respectively. This roughly correspond to a 2.5-year period. The reason for the choice of such a long time period is technical complications with numerical convergence for smaller values of δ , and future versions of this work will address this technical issue.

Simulations confirm qualitative approximation results above for price and quantity dynamics. We start by exploring the dynamic path towards the steady state when both cost and demand shocks are switched off ($W_t \equiv \zeta_t \equiv 1$), and the initial stock of the durable good is zero ($D_{-1} = 0$). Figure 3.1 reports the path of prices in the

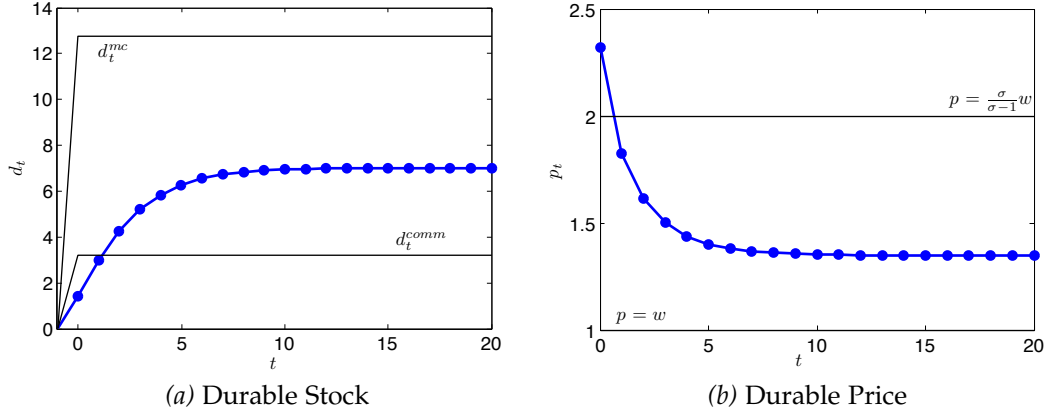


Figure 3.1: Equilibrium dynamic path of quantities and prices, no shocks

right panel and the path of quantities in the left panel. We contrast the equilibrium dynamics without commitment with the case of marginal-cost pricing and monopoly-pricing with commitment. In both of these benchmark cases there is no dynamics in prices and durable stock jumps instantaneously to its steady state level. The dynamics without commitment is substantially different: prices gradually decline as stock gradually rises, reaching steady state only after a number of periods. This illustrates endogenous dynamics in this case and the role of the durable stock as the state variable.

Next, in Figures 3.2–3.3, we consider the response of monopolist price, markup and durable stock to an unanticipated permanent cost and demand shocks. Confirming our theoretical findings, markup is decreasing in response to a cost rise, and increasing in response to a positive demand shock. This endogenous response of markup and price dampens the short-run response of quantities to both shocks. As a result, quantities adjust only gradually to both shocks, which is contrasted with the immediate adjustment of quantities under both marginal-cost and constant-markup

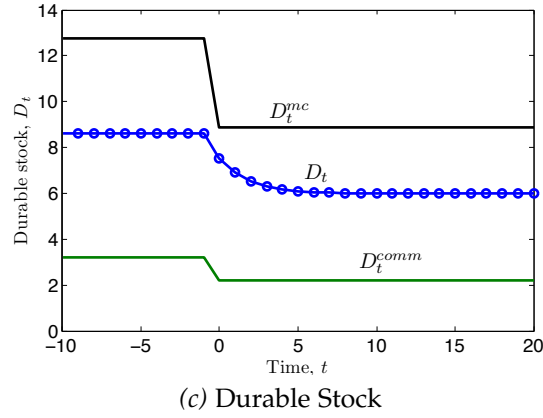
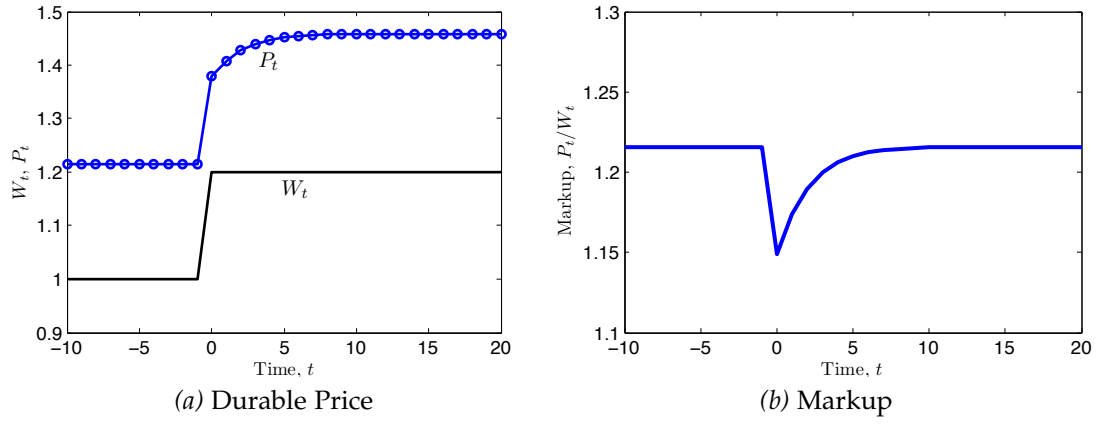


Figure 3.2: Response to a one-time cost shock

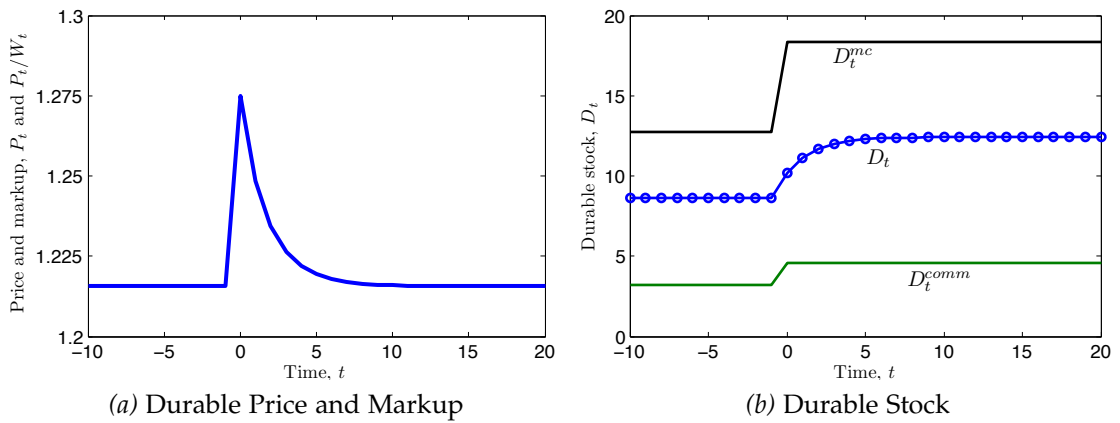


Figure 3.3: Response to a one-time demand shock

(the case of commitment) pricing.

Table 3.1: Equilibrium Dynamics under Cost Shocks

| $\log(\cdot)$ | σ (%) | ρ | $\text{corr}(\cdot, \log W_t)$ |
|--------------------------|--------------|--------|--------------------------------|
| Wage, W_t | 4.9 | 0.80 | 1.00 |
| Price, P_t | 5.1 | 0.90 | 0.88 |
| Markup, P_t/W_t | 2.2 | 0.69 | -0.19 |
| Durable stock, D_t | | | |
| — constant markup | 15.5 | 0.79 | -0.99 |
| — discretion | 12.2 | 0.95 | -0.75 |
| — ratio (disc/comm) | | | 0.29 |
| Durable purchases, X_t | | | |
| — constant markup | 70.7 | -0.08 | -0.31 |
| — discretion | 21.4 | 0.57 | -0.91 |
| — ratio (disc/comm) | | | 0.16 |

Finally, we simulate the partial equilibrium dynamics under stochastic cost and demand shocks. First, we consider stochastic cost shocks evolving according to a discretized version of an $AR(1)$ with standard deviation of innovation of 5% and a persistence of 0.8. Table 3.1 reports the statistical properties prices and quantities in this dynamic equilibrium. Markup is indeed countercyclical and price exhibits endogenous persistence, in excess of that of the exogenous cost process. Furthermore, durable good purchases are mildly negatively correlated under constant-markup pricing, while they become strongly positively autocorrelated under solution without commitment. This is an empirically appealing property of durable pricing without commitment, which allows to obtain realistic dynamics of durable purchases even in the absence of adjustment costs. Lastly, endogenous markup variation without commitment substantially reduces the volatility of both durable purchases and durable stock.

Table 3.2 reports the pass-through coefficients from the regression of prices on costs, with and without the lagged cost variable. The upper panel runs the pass-

Table 3.2: Pass-through of Cost Shocks

| | $\log W_t$ | $\log W_{t-1}$ |
|-------------------|-------------------|-----------------------|
| $\log P_t$ | 0.91 | |
| $\log P_t$ | 0.65 | 0.34 |
| | $\Delta \log W_t$ | $\Delta \log W_{t-1}$ |
| $\Delta \log P_t$ | 0.61 | |
| $\Delta \log P_t$ | 0.63 | 0.15 |

through regression in levels, while the lower panel produces results in differences. Pass-through is incomplete (91% in levels and 61% in differences) with over 2/3 of pass-through happening on impact and the remaining pass-through after one period.

Table 3.3: Equilibrium Dynamics under Demand Shocks

| $\log(\cdot)$ | σ (%) | ρ | $\text{corr}(\cdot, \log \tilde{\zeta}_t)$ |
|---------------------------|--------------|--------|--|
| Demand, $\tilde{\zeta}_t$ | 4.8 | 0.77 | 1.00 |
| Price and markup, P_t/W | 1.9 | 0.79 | 0.18 |
| Durable stock, D_t | | | |
| — constant markup | 9.7 | 0.77 | 1.00 |
| — discretion | 7.2 | 0.94 | 0.66 |
| — ratio (disc/comm) | | | -0.22 |
| Durable purchases, X_t | | | |
| — constant markup | 36.1 | -0.03 | 0.91 |
| — discretion | 13.6 | 0.56 | 0.55 |

Finally, we consider the dynamic equilibrium with demand shocks. We assume that demand shocks follow a discretized $AR(1)$ process also with standard deviation of innovation of 5% and persistence of 0.8. Table 3.3 reports the statistical properties of the dynamics of prices and quantities in this stochastic equilibrium. The markup is procyclical in this case, with a standard deviation of about 2% and persistence of about 0.8. This again results in positively autocorrelated durable purchases in contrast with *iid* durable purchases under marginal-cost or constant-markup pricing. In addition, both durable purchases and durable stock is less volatile under pricing

without commitment, as in the case of cost shocks.

3.3.3 Extension: Calvo sticky price setting

Result 7 (Equivalence of Calvo and Flexible Pricing): The solution of a durable-good monopolist pricing problem under Calvo price stickiness with probability of price non-adjustment θ is identical to the problem of a flexible price durable-good monopolist under the following substitution of primitives: $\beta \rightarrow \beta_C$, $\delta \rightarrow \delta_C$, $u \rightarrow u_C$, where

$$\begin{aligned}\beta_C &\equiv \beta \frac{1-\theta}{1-\beta\theta}, \\ \delta_C &\equiv 1 - \frac{1-\beta\theta}{1-\beta\theta(1-\delta)}(1-\delta), \\ u_C(D) &\equiv \frac{1}{1-\beta\theta(1-\delta)}u(D).\end{aligned}\tag{3.17}$$

Proof: We have seen that with flexible prices the dynamics can be summarized as

$$V(D_-) = \max_P \{ (P - W) (d(P) - (1 - \delta)D_-) + \beta V(d(P)) \}, \tag{3.18}$$

which yields policy function $p(D_-)$, and where the consumer reaction function $d(P)$ is implicitly defined by the consumer first order condition

$$u'(d(P)) = P - \beta(1 - \delta)p(d(P)). \tag{3.19}$$

To keep notation simple, we suppressed the time indices, replacing (P_t, D_{t-1}) by (P, D_-) .

Now let us consider the case of Calvo price setting. Denote by $1 - \theta$ the probability that the firm is allowed to adjust prices in any definite period. The consumer first

order condition will become

$$u'(d(P)) = P - \beta(1 - \delta)(\theta P + (1 - \theta)p(d(P))),$$

where p is again the firm's desired price as the function of the state variable. Using definitions (3.17), this may be rewritten as

$$u'_C(d(P)) = P - \beta_C(1 - \delta_C)p(d(P)). \quad (3.20)$$

Note that demand as a function of P is the same irrespective of whether the firm adjusted the price this period: consumers take the current price as given and the expected price next period (for a given stock of durables) does not depend on whether the firm adjusted the price in the current period.

The problem of the firm in a period of price adjustment is

$$V(D_-) = \max_P \{ (P - W)(d(P) - (1 - \delta)D_-) + \beta(1 - \theta)V(d(P)) + \beta\theta U(P) \},$$

where $U(P)$ is the value of the firm if the price is not adjusted; i.e., $U(P)$ is given by

$$U(P) = (P - W)\delta d(P) + \beta(1 - \theta)V(d(P)) + \beta\theta U(P).$$

Note that if the firm does not adjust prices, consumers will consume the same amount as in the previous period. Combining the two expressions, leads to the Bellman equation

$$V(D_-) = \max_P \left\{ (P - W) \left[\frac{1 - \beta(1 - \delta)\theta}{1 - \beta\theta} d(P) - (1 - \delta)D_- \right] + \frac{\beta(1 - \theta)}{1 - \beta\theta} V(d(P)) \right\}.$$

Recalling the definitions (3.17), we see that problem is equivalent to

$$\tilde{V}(D_-) = \max_P \{ (P - W) (d(P) - (1 - \delta_C)D_-) + \beta_C \tilde{V}(d(P)) \}, \quad (3.21)$$

with

$$\tilde{V}(D) \equiv \frac{1 - \beta\theta}{1 - \beta(1 - \delta)\theta} V(D).$$

Equations (3.20) and (3.21) are the same as (3.19) and (3.18) with the replacement $\beta \rightarrow \beta_C$, $\delta \rightarrow \delta_C$, $u \rightarrow u_C$, $V(D) \rightarrow \tilde{V}(D)$. This implies the equivalence of the dynamics with flexible prices and with Calvo price setting stated above. ■

3.4 Durable good oligopoly

We have analyzed the price dynamics of durable goods in the case of monopoly. In this section, we will investigate the case of Cournot competition of N symmetric oligopolistic firms producing a homogenous durable good. The firms are unable to commit to future production policy. The production decision of each firm is influenced by two different forces: competition with the other firms and inter-temporal competition with itself, since the future decision makers at the firm do not have the same objectives. Macroeconomic literature has focused on the first effect, assuming that for firms production decisions are based on the current competition, see (e.g., Barsky, House, and Kimball, 2007). Here we model both forces jointly. As we will see, both of them push prices closer to marginal cost.

The stock of the durable good is the only state variable. For simplicity, we do not consider stochastic cost shocks and demand shocks here, although the analysis can be extended to parallel our discussion of monopoly. In this simpler setting, one can still analyze the endogenous price dynamics in response to permanent shocks and

identify the value of the persistence parameter. The following lemma characterizes the price dynamics in the case under consideration.

Lemma 8 (Durable-good oligopoly equilibrium dynamics): The equilibrium in the durable-good Cournot oligopoly market with N symmetric firms with constant marginal cost W is described by the following system:

$$\begin{aligned} (P_t - W) - \beta(P_{t+1} - W) & \left(\frac{1 - \delta}{N} + \frac{N - 1}{N} f'(D_t) \right) \\ & = - \frac{D_t - (1 - \delta)D_{t-1}}{N} (u''(D_t) + \beta(1 - \delta)p'(D_t)) \end{aligned} \quad (3.22)$$

$$u'(D_t) = P_t - \beta(1 - \delta)P_{t+1}, \quad (3.23)$$

where

$$D_t = f(D_{t-1}), \quad P_t = p(D_{t-1}).$$

Proof: Consumer optimization immediately implies

$$P_t = u'(D_t) + \beta(1 - \delta)p(D_t), \quad (3.24)$$

where $p(D_t)$ describes the dependence of the price of the durable good in the next period on the stock of the good in the current period. This is the second equation of the lemma. Denote by $x(D_{t+1})$ the equilibrium strategy (i.e., the amount produced) of every oligopolist a function of the state variable. Suppose that in period t one firm deviates from its equilibrium strategy $x(D_{t-1})$, and produces \tilde{x}_t instead. In this case

$$D_t = (1 - \delta)D_{t-1} + \tilde{X}_t + (N - 1)x(D_{t-1}). \quad (3.25)$$

Since the firm is free to choose its optimal level of production, its value satisfies the Bellman equation

$$V(D_{t-1}) = \max_{\tilde{X}_t} \{ (P_t(\tilde{x}_t) - W) \tilde{X}_t + \beta V(D_t(\tilde{X}_t)) \},$$

where the dependence of P_t and D_t on \tilde{X}_t is given by equations (3.24) and (3.25). Denote by λ and μ the Lagrange multipliers on those two equations. The envelope condition and the first order conditions corresponding to the choice of \tilde{X}_t , P_t and D_t may be written as

$$\begin{aligned} V'(D_{t-1}) &= -\mu \left((1 - \delta) + (N - 1)x'(D_{t-1}) \right), \\ P_t - W &= \mu, \\ \beta V'(D_t) + \mu &= -\lambda \left(u''(D_t) + \beta(1 - \delta)p'(D_t) \right), \\ \tilde{X}_t &= \lambda. \end{aligned}$$

Substituting out the Lagrange multipliers and imposing the equilibrium requirement $\tilde{X}_t = x(D_{t-1})$ yields

$$\begin{aligned} & (P_t - W) - \beta(P_{t+1} - W) \left((1 - \delta) + (N - 1)x'(D_t) \right) \\ &= -x(D_{t-1}) \left(u''(D_t) + \beta(1 - \delta)p'(D_t) \right) \end{aligned}$$

Just like (3.8), this condition can be justified by a simple perturbation argument (see the discussion following Lemma 2). The firm sells a little bit more in period t and a little less in period $t + 1$, with this amounts chosen in a way that lead to the original equilibrium path from period $t + 2$ onward. In choosing the quantity deviations for the perturbation argument, one needs to take into account the reaction of the

competitors to the deviation in period t , reflected by the presence of $(N - 1)x'(D_t)$ on the left hand side.

This condition together the following definition of the function f

$$D_t = f(D_{t-1}) = (1 - \delta)D_{t-1} + Nx(D_{t-1}), \text{ with } f'(D_{t-1}) = (1 - \delta) + Nx'(D_{t-1}).$$

implies the first equation of the lemma. ■

For $N = 1$, the system of equations in the lemma agrees with the monopoly results. For $N = 1$ we observe the following two changes to the system of equations. First, on the right hand side instead of $X_t = D_t - (1 - \delta)D_{t-1}$ we have X_t/N , which implies that the profit losses of an oligopolist from lower price are shared equally with the $(N - 1)$ competitors. This force is present also in the case of nondurable goods ($\delta = 1$), and it reduces the price charged by the monopolist. Second, in the spirit of the perturbation argument, if a firm sells one additional unit of the durable good in period t , and wants to compensate for this in period $t + 1$ to return to the original equilibrium path, its sales should not drop by $(1 - \delta)$, but by $(1 - \delta)/N$ plus the change in sales by its competitors, reflected by $x'(D_t)(N - 1)/N$. Since $x'(D_t) < 0$, this is less than $(1 - \delta)$.

Lemma 9 (Oligopoly with quadratic utility): With N symmetric firms, constant marginal cost W , and quadratic utility $u'(D_t) = a - bD_t$, there exists a 'linear' equilibrium of the form:

$$D_t = \bar{D} + \phi(D_{t-1} - \bar{D}) \tag{3.26}$$

$$P_t = \bar{P} - \alpha(D_{t-1} - \bar{D}) \tag{3.27}$$

which solves the equilibrium system of equations exactly. The positive persistence parameter ϕ satisfies the cubic⁴ equation

$$\beta (N - 1) \phi^3 + \beta (1 - \delta) \phi^2 - (N + 1) \phi + 1 - \delta = 0. \quad (3.28)$$

With the knowledge of ϕ , the parameters α , \bar{D} , and \bar{P} may be recovered as

$$\begin{aligned} \alpha &= \frac{b\phi}{\gamma_2} \\ \bar{D} &= \frac{1}{b} \left(\frac{a}{\gamma_1} - W \right) \frac{\gamma_1 + (N-1)\gamma_2}{\gamma_1 + (N-1)\gamma_2 + \frac{\delta}{\gamma_2}} \\ \bar{P} &= \frac{a}{\gamma_1} - \frac{1}{\gamma_1} \left(\frac{a}{\gamma_1} - W \right) \frac{\gamma_1 + (N-1)\gamma_2}{\gamma_1 + (N-1)\gamma_2 + \frac{\delta}{\gamma_2}} \end{aligned}$$

where

$$\gamma_1 \equiv 1 - \beta (1 - \delta), \quad \gamma_2 \equiv 1 - \beta (1 - \delta) \phi$$

Proof: Plugging the assumed form of the solution (3.26) and (3.27) into the dynamic equations (3.22) and (3.23) and comparing coefficients of different terms leads to the four conditions

$$\begin{aligned} \frac{\alpha}{\phi} \frac{\delta \bar{D}}{\bar{P} - W} &= 1 - \beta (1 - \delta) + (N - 1) (1 - \beta \phi) \\ 0 &= 1 - \frac{1 - \delta}{\phi} + N - \beta \phi (1 - \delta + (N - 1) \phi) \\ u'(\bar{D}) &= (1 - \beta (1 - \delta)) \bar{P} \\ -u''(\bar{D}) \frac{\phi}{\alpha} &= 1 - \beta (1 - \delta) \phi \end{aligned}$$

Here we used $p'(D_{t-1}) = -\alpha$, as implied by (3.27). The second condition may be rewritten as (3.28). The remaining three conditions can be manipulated to give the explicit expressions for α , \bar{D} , and \bar{P} in the lemma. It is straightforward to check that with these values of the parameters, the dynamic equations (3.22) and (3.23) are

⁴ Of course, any cubic equation can be solved algebraically. For brevity, we will omit the explicit expressions for the solution here.

identically satisfied. ■

Result 10 *Both ϕ and α decrease in N .*

Proof: Treat the parameter N in (3.28) as a continuous variable. Implicit function theorem then implies

$$-\frac{d\phi}{dN} = \frac{\phi - \beta\phi^3}{N + 1 - 2\beta(1 - \delta)\phi - 3\beta(N - 1)\phi^2}$$

Since $\phi < 1$, the numerator is clearly positive. Algebraic manipulation may be used to show that the denominator is also positive. This leads to the conclusion that ϕ is a decreasing function of N . The explicit expression for α is $\alpha = b\phi / (1 - \beta(1 - \delta)\phi)$. This is obviously an increasing function of ϕ , which implies that α decreases in response to increased N . ■

3.5 Durable good monopolistic competition

We now consider the case of monopolistic competition with infinitesimal firms producing imperfectly substitutable varieties. The firms cannot commit to a path of future prices. We specialize to the cases where the influence of other firms on the demand for a particular variety can be summarized by a simple sufficient statistic.

Consider the case of a continuum of varieties of durable goods, distinguished by index $i \in \Omega \equiv [0, 1]$. The varieties depreciate at the same rate δ , and the stock of variety i evolves according to

$$D_{it} = X_{it} + (1 - \delta)D_{i,t-1},$$

where X_{it} is quantity of variety i sold in t . Denote the price of variety i by P_{it} . The

consumer maximizes

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t U(C_t, \{D_{it}\}_{i \in \Omega})$$

subject to

$$\sum_{t=0}^{\infty} \beta^t \left(P_{C,t} C_t + \int_{i \in \Omega} P_{it} (D_{it} - (1 - \delta) D_{i,t-1}) di \right) = NPV_0$$

with $D_{i,-1}$ and NPV_0 given⁵. The consumer's first order condition is

$$\frac{\delta U(C_t, \{D_{it}\}_{i \in \Omega})}{\delta D_{it}} = \lambda_t P_{it} - \beta(1 - \delta) \mathbb{E}_t \lambda_{t+1} P_{i,t+1},$$

where $\delta U(C_t, \{D_{it}\}_{i \in \Omega}) / \delta D_{it}$ is the functional derivative of $U(C_t, \{D_{it}\}_{i \in \Omega})$ with respect to D_{it} , and $\lambda_t \equiv U_{1,t} / P_{C,t}$. Consider the case of quasi-linear utility function $U(C_t, \{D_{it}\}_{i \in \Omega}) = C_t + u(\{D_{it}\}_{i \in \Omega})$. The price of the nondurable good $P_{C,t}$ will be normalized to one, implying $\lambda_t = 1$. The first order condition becomes

$$\frac{\delta u(\{D_{it}\}_{i \in \Omega})}{\delta D_{it}} = P_{it} - \beta(1 - \delta) \mathbb{E}_t P_{i,t+1}.$$

Let us specialize to cases⁶ where the influence of competitors on firm i may be summarized by a sufficient statistic $D_t \equiv \Psi(\{D_{it}\}_{i \in \Omega})$ in the sense that

$$\frac{\delta u(\{D_{it}\}_{i \in \Omega})}{\delta D_{it}} = v(D_{it}; D_t).$$

The demand equation now takes the form

$$v(D_{it}; D_t) = P_{it} - \beta(1 - \delta) \mathbb{E}_t P_{i,t+1}.$$

⁵ Generalization to consumer's stochastic endowment is straightforward.

⁶ These include CES utility functions, as well as quadratic utility functions discussed below.

Result 11 *For a given process D_t , the problem of durable good monopolistic competitor described above is equivalent to the problem of a durable good monopolist with stochastic demand shocks.*

For simplicity, let us consider the case without uncertainty: $W_t = W$. The following lemma characterizes the price dynamics.

Lemma 12 (Monopolistic competition, pricing without commitment): The equilibrium prices and quantities satisfy

$$(P_{it} - W) - \beta(1 - \delta)(P_{i,t+1} - W) \quad (3.29)$$

$$= (D_{it} - (1 - \delta)D_{i,t-1}) (v'(D_{it}; D_t) + \beta(1 - \delta)p'(D_{it}; D_t)),$$

$$v(D_t; D_{it}) = P_{it} - \beta(1 - \delta)P_{i,t+1} \quad (3.30)$$

where $D_t \equiv \Psi(\{D_{it}\}_{i \in \Omega})$, and the policy function $p(D_{it}; D_t)$ describes the dependence of $P_{i,t+1}$ on D_{it} and D_t .

Proof: Equation (3.30) has been derived above. Equation (3.29) is an immediate consequence of Lemma 2, with the appropriate renaming of variables. This is because of the equivalence of the problem of durable good monopolistic competitor and durable good monopolist facing shifts in demand. ■

Consider now the case of quadratic utility. Our goal is to identify the persistence parameters corresponding to deviations from the steady state. The utility will be parameterized as

$$u(\{D_{it}\}_{i \in \Omega}) = a \int_{i \in \Omega} D_{it} di - \frac{b_1}{2} \int_{i \in \Omega} D_{it}^2 di - \frac{b_2}{2} \left(\int_{i \in \Omega} D_{it} di \right)^2$$

This type of utility function has been used, for example, in Ottaviano, Tabuchi, and Thisse (2002) and Melitz and Ottaviano (2008). The function $v(D_{it}; D_t)$ now takes the form

$$v(D_{it}; D_t) = a - b_1 D_{it} - b_2 D_t \quad (3.31)$$

with

$$D_t \equiv \int_{i \in \Omega} D_{it} di$$

being the average stock of durable goods at time t . The following lemma characterizes the equilibrium path

Lemma 13 (Monopolistic competition, quadratic utility): The evolution of prices and quantities is described by

$$D_{it} = \bar{D} + \phi_\kappa(D_{t-1} - \bar{D}) + \phi_\alpha(D_{i,t-1} - D_{t-1}), \quad (3.32)$$

$$P_{it} = \bar{P} - \kappa(D_{t-1} - \bar{D}) - \alpha(D_{i,t-1} - D_{t-1}), \quad (3.33)$$

$$D_t \equiv \int_{i \in \Omega} D_{it} di, \quad P_t \equiv \int_{i \in \Omega} P_{it} di.$$

The parameters are given by

$$\begin{aligned}
\phi_\alpha &= \frac{1 - \sqrt{1 - \beta(1 - \delta)^2}}{\beta(1 - \delta)}, \\
\alpha &= \frac{b_1}{\beta(1 - \delta)} \left(\frac{1}{\sqrt{1 - \beta(1 - \delta)^2}} - 1 \right), \\
\phi_\kappa &= \frac{1 - \delta}{1 + \frac{b_1 + b_2}{b_1} \sqrt{1 - \beta(1 - \delta)^2}}, \\
\kappa &= \frac{(1 - \delta)(b_1 + b_2)}{1 - \beta(1 - \delta)^2 + \frac{b_1 + b_2}{b_1} \sqrt{1 - \beta(1 - \delta)^2}}, \\
D_s &= \frac{a - (1 - \beta(1 - \delta))W}{\frac{\delta b_1}{\sqrt{1 - \beta(1 - \delta)^2}} + b_1 + b_2}, \\
P_s &= \frac{1}{1 + \frac{b_1 + b_2}{\delta b_1} \sqrt{1 - \beta(1 - \delta)^2}} \frac{a}{1 - \beta(1 - \delta)} + \frac{\frac{\delta b_1}{b_1 + b_2} \frac{1}{\sqrt{1 - \beta(1 - \delta)^2}} + 1}{1} W.
\end{aligned} \tag{3.34}$$

Proof: With $v(D_{it}; D_t)$ taking the form (3.31), the dynamic equations (3.29) and (3.30) become

$$(P_{it} - W) - \beta(1 - \delta)(P_{i,t+1} - W) = (D_{it} - (1 - \delta)D_{i,t-1})(-b_1 + \beta(1 - \delta)p'(D_{it}; D_t))$$

$$a - b_1 D_{it} - b_2 D_t = P_{it} - \beta(1 - \delta)P_{i,t+1}$$

Plugging the assumed form of the equilibrium (3.32) and (3.33) into these equations and comparing coefficients of different terms leads to six conditions:

$$\begin{aligned}
(1 - \beta(1 - \delta))\bar{P} &= a - (b_1 + b_2)\bar{D} \\
1 - \beta(1 - \delta)\phi_\kappa &= (b_1 + b_2)\frac{\phi_\kappa}{\kappa} \\
1 - \beta(1 - \delta)\phi_\alpha &= b_1\frac{\phi_\alpha}{\alpha} \\
\frac{\alpha}{\phi_\alpha}\delta\bar{D} &= (1 - \beta(1 - \delta))(\bar{P} - W) \\
-\frac{\phi_\kappa\alpha}{\phi_\alpha\kappa}\left(1 - \frac{1 - \delta}{\phi_\kappa}\right) &= 1 - \beta(1 - \delta)\phi_\kappa \\
-\left(1 - \frac{1 - \delta}{\phi_\alpha}\right) &= 1 - \beta(1 - \delta)\phi_\alpha
\end{aligned}$$

Note that here we used $p'(D_{i,t-1}; D_{t-1}) = -\alpha$, implied by (3.33). Algebraic manipulations lead the result (3.34). It is straightforward to verify that with these values all equilibrium conditions are satisfied. ■

Result 14 *The collective persistence parameter is smaller than the idiosyncratic persistence parameter: $\phi_\kappa < \phi_\alpha$.*

Proof: The values of these parameters given in Lemma 13 may be rewritten as

$$\phi_\kappa = \frac{1 - \delta}{1 + \frac{b_1 + b_2}{b_1} \sqrt{1 - \beta(1 - \delta)^2}}, \quad \phi_\alpha = \frac{1 - \delta}{1 + \sqrt{1 - \beta(1 - \delta)^2}}.$$

Since $(b_1 + b_2) / b_1 > 1$, the inequality $\phi_\kappa < \phi_\alpha$ immediately follows. ■

Result 15 *The price P_{it} of a particular variety of durable goods depends more strongly on the deviations of the average stock of durables D_{t-1} from the steady state value \bar{D} than on the deviations of the stock $D_{i,t-1}$ of this variety from the average D_{t-1} : $\kappa > \alpha$.*

Proof: Start with the inequality

$$\frac{1}{\frac{b_1}{b_1 + b_2} (1 - \beta(1 - \delta)^2) + \sqrt{1 - \beta(1 - \delta)^2}} > \frac{1}{1 - \beta(1 - \delta)^2 + \sqrt{1 - \beta(1 - \delta)^2}},$$

which is certainly satisfied due to $(b_1 + b_2) / b_1 > 1$. Multiplying both sides by $(1 - \delta) b_1$ and then manipulation each side separately leads to

$$\frac{(1 - \delta)(b_1 + b_2)}{1 - \beta(1 - \delta)^2 + \frac{b_1 + b_2}{b_1} \sqrt{1 - \beta(1 - \delta)^2}} > \frac{b_1}{\beta(1 - \delta)} \left(\frac{1}{\sqrt{1 - \beta(1 - \delta)^2}} - 1 \right),$$

which is, due to Lemma 13, equivalent to $\kappa > \alpha$. ■

3.6 *Conclusion*

We evaluate price and quantity dynamics in several environments such as monopoly, oligopoly and monopolistic competition. We show that in all these environment, in response to cost shocks, markups are countercyclical and therefore pass-through is incomplete. We contrast these findings with that of nondurable goods pricing results. We also contrast this to the case of marginal cost pricing of durable goods which is the typical assumption in the macroeconomic literature.

In this work we have limited our study to the pricing problem of firms and therefore the analysis has been partial equilibrium in nature. In future versions we plan to add general equilibrium elements into the model and study the implications of the pricing dynamics of durable goods for aggregate macroeconomic variables both in closed and open economy environments.

APPENDIX

A. APPENDIX TO CHAPTER 1

A.1 Neglecting changes in general equilibrium effects

Consider the Krugman (1980) model in the case of a completely symmetric circle, as in Fig. 1.1a. Solving for the equilibrium is simple because the GDP density will be the same everywhere. Now suppose we would like to know the response to changes in trade costs. To be concrete, let us split the circle into two semicircular ‘countries’, and introduce additional ‘iceberg’ type border costs, as in Anderson and van Wincoop (2003), i.e. as goods cross the border a certain fixed percentage is lost.

The consequences of this change in trade costs are illustrated in Fig. A.1b, which captures all general equilibrium effects. At no location will the GDP increase when the border costs are introduced. If we decided to neglect general equilibrium feedback, in the sense of neglecting the first term on the right-hand-side of (1.5) or (1.18), the calculations would be much simpler, and we would get Fig. A.1a. The results are quite different. In certain regions we would not get even the sign of the overall effect right.

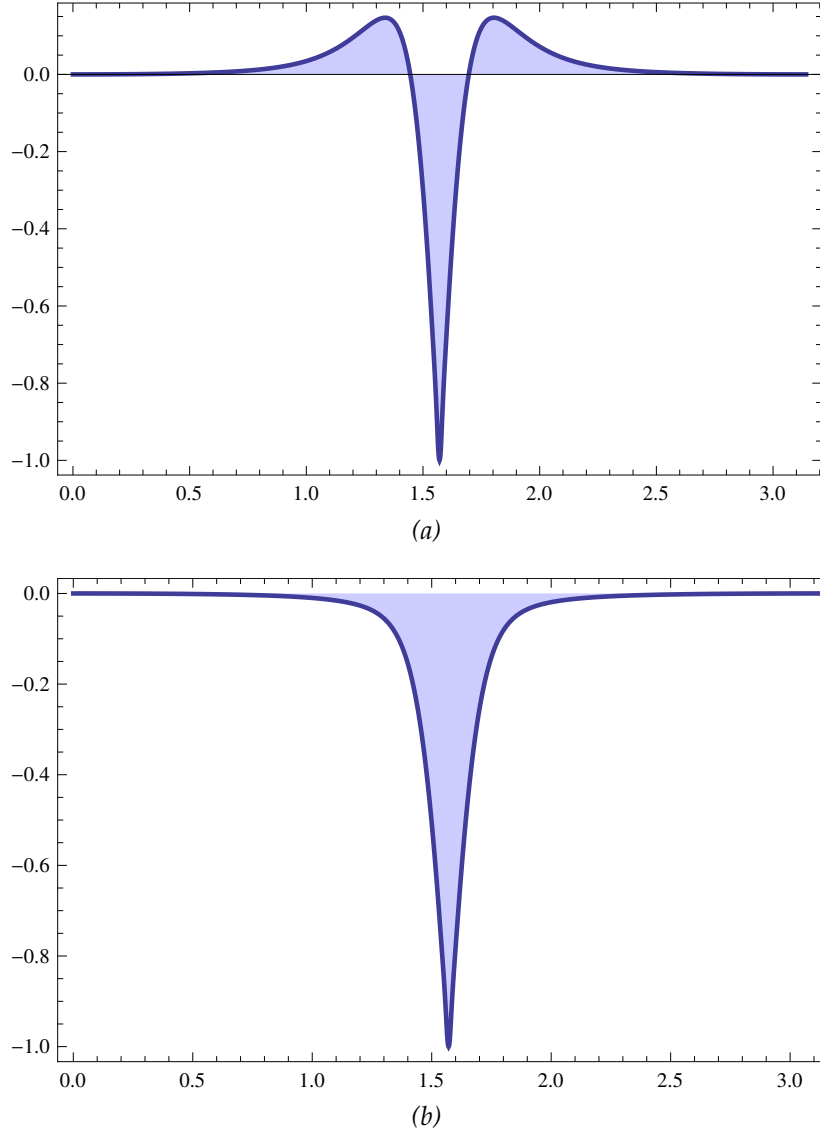


Figure A.1: The figure shows the first-order response of GDP to increased border costs. The spatial configuration is a circle parameterized by $\theta \in (-\pi, \pi]$, with only the range $[0, \pi]$ shown in the figure, which is sufficient due to the left-right symmetry. The circle is split into two semi-circular countries with the two border points located at $\theta = \pm\pi/2$. Part (a) plots the first-order change in GDP induced by increasing border costs as calculated ignoring general equilibrium feedback, while part (b) presents the full general equilibrium result. The parameter values used to generate the figure are $\sigma = 6$, $\rho = 0.75$, and $\alpha R = 5$, and the functional form of trade costs is $(1 + 4\alpha^2 R^2 \sin^2(\theta/2))^{\rho/2}$. For simple comparison, the y-axes are linearly transformed.

A.2 *Remarks on methodology: reverse engineering equilibria from comparative statics*

A.2.1 *The case of a single endogenous variable*

International trade models are fairly complicated. In order to make the general computational strategy employed in the present work easier to follow, this appendix illustrates some intuition used extensively in the main text with elementary examples, not necessarily coming from trade theory. Readers who find the rest of the present work intuitively clear may prefer to skip this discussion, as it does not contain any novel economic insights.

Consider an economic model in which the equilibrium value of an endogenous variable y is given implicitly as a function of an exogenous parameter κ by the equation

$$f(y, \kappa) = 0.$$

where the known function f satisfies the requirements of the implicit function theorem. For example, one can think of $f(y, \kappa) = 0$ as representing the first-order condition of a maximization problem. Suppose that we know the value y_0 corresponding to $\kappa = 0$, i.e. $f(y_0, 0) = 0$. Assume also that it is possible to compute all partial derivatives of f at $(y, \kappa) = (y_0, 0)$. It may be that the function $f(y, \kappa)$ is hard to invert with respect to its first argument. Under such circumstances, we can still recover the solution to the economic problem $y(\kappa)$ using comparative statics, assuming that $y(\kappa)$ is an analytic function.

First of all, since the first partial derivatives are known, we can use the approxi-

mation

$$y(\kappa) = y_0 + \left. \frac{dy}{d\kappa} \right|_{\kappa=0} \kappa + O(\kappa^2),$$

where the derivative may be computed as

$$\left. \frac{dy}{d\kappa} \right|_{\kappa=0} = -\frac{f_2(y_0, 0)}{f_1(y_0, 0)}. \quad (\text{A.1})$$

This is what first-order comparative statics teaches us. But of course, we can go further. With higher precision,

$$y(\kappa) = y_0 + \left. \frac{dy}{d\kappa} \right|_{\kappa=0} \kappa + \frac{1}{2} \left. \frac{d^2y}{d\kappa^2} \right|_{\kappa=0} \kappa^2 + O(\kappa^3).$$

The second derivative here can be obtained by the standard formula for second-order comparative statics,

$$\left. \frac{d^2y}{d\kappa^2} \right|_{\kappa=0} = - \left. \frac{f_2^2 f_{11} - 2f_1 f_2 f_{12} + f_1^2 f_{22}}{f_1^3} \right|_{y=y_0, \kappa=0}. \quad (\text{A.2})$$

In principle we can evaluate any derivative $d^n y / d\kappa^n$, and recover the full solution to the economic problem as the series

$$y(\kappa) = y_0 + y_1 \kappa + y_2 \kappa^2 + y_3 \kappa^3 + \dots$$

with

$$y_n \equiv \frac{1}{n!} \left. \frac{d^n y}{d\kappa^n} \right|_{\kappa=0}.$$

This is of course not as elegant as inverting $f(y, \kappa)$ with respect to its first argument directly, but it conveys the same information.

Obviously, we need a systematic way to express y_n in terms of partial derivatives

of f . But that is not difficult. Substituting the Taylor expansion of $y(\kappa)$ for y into $f(y, \kappa) = 0$, we get

$$f(y_0 + y_1\kappa + y_2\kappa^2 + y_3\kappa^3 + \dots, \kappa) = 0.$$

The Taylor series of the left hand side is

$$\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{f^{(j,k)}(y_0, 0)}{j!k!} \left(\sum_{l=1}^{\infty} y_l \kappa^l \right)^j \kappa^k,$$

where $f^{(j,k)}$ is the j th partial derivative with respect to the first argument of f and the k th partial derivative with respect to the second argument. The relation must hold for any κ , so for any non-negative integer n , the term proportional to κ^n must vanish. For $n = 0$, this implies $f(y_0, 0) = 0$, which is satisfied by assumption. For $n = 1$, the requirement becomes

$$f_1(y_0, 0) y_1 \kappa + f_2(y_0, 0) \kappa = 0,$$

which is equivalent to the first-order comparative statics (A.1). For $n = 2$,

$$f_1(y_0, 0) y_2 \kappa^2 + \frac{1}{2} f_{11}(y_0, 0) y_1^2 \kappa^2 + f_{12}(y_0, 0) y_1 \kappa^2 + \frac{1}{2} f_{22}(y_0, 0) \kappa^2 = 0,$$

leading to the second-order comparative statics formula (A.2). The value of y_3 can be obtained by looking at terms proportional to κ^3 , etc.

Of course, in concrete applications of this method, it is very likely that most economic intuition is already contained in the first few terms, say y_0, y_1 , and y_2 . Only under rare circumstances would one need to compute even higher terms. This makes the present approach useful at the practical level.

A.2.2 The case of two endogenous variables and its generalization

The case of a single endogenous variable was straightforward. Now suppose that instead, y is a two-dimensional vector, $y \equiv (y_{(x_1)}, y_{(x_2)})'$, where the labels x_1 and x_2 can be thought of as two distinct locations in space. Let the equations following from the model take the implicit form,

$$f(y, \kappa) = 0,$$

where $f(y, \kappa)$ is now a two-dimensional vector as well. Assume also that the first component of this equation can be solved with respect to $y_{(x_1)}$ and the second one with respect to $y_{(x_2)}$, i.e. that for some known functions $g_{(x_1)}$ and $g_{(x_2)}$, the equations

$$y_{(x_1)} = g_{(x_1)}(y_{(x_2)}, \kappa),$$

$$y_{(x_2)} = g_{(x_2)}(y_{(x_1)}, \kappa)$$

are equivalent to $f(y, \kappa) = 0$. This assumption is made for expositional purposes only, and can be easily lifted. The task is again the same as in the single endogenous variable case. We are given the solution to these equations $y_0 = (y_{0(x_1)}, y_{0(x_2)})'$, corresponding to $\kappa = 0$, and need to find y for non-zero κ , or at least the first-order change y_1 defined by $y = y_0 + y_1\kappa + O(\kappa^2)$.

One intuitive way to approach the problem is the following. Denote

$$v \equiv \begin{pmatrix} v_{(x_1)} \\ v_{(x_2)} \end{pmatrix} = \begin{pmatrix} \left. \frac{\partial g_{(x_1)}}{\partial \kappa} \right|_{y_{(x_2)}=y_{0(x_2)}, \kappa=0} \\ \left. \frac{\partial g_{(x_2)}}{\partial \kappa} \right|_{y_{(x_1)}=y_{0(x_1)}, \kappa=0} \end{pmatrix},$$

$$G \equiv \begin{pmatrix} G_{(x_1, x_1)} & G_{(x_1, x_2)} \\ G_{(x_2, x_1)} & G_{(x_2, x_2)} \end{pmatrix} \equiv \begin{pmatrix} 0 & \left. \frac{\partial g_{(x_1)}}{\partial y_{(x_2)}} \right|_{y_{(x_2)}=y_{0(x_2)}, \kappa=0} \\ \left. \frac{\partial g_{(x_2)}}{\partial y_{(x_1)}} \right|_{y_{(x_1)}=y_{0(x_1)}, \kappa=0} & 0 \end{pmatrix}.$$

If the exogenous parameter changes from 0 to some small value κ , it is natural to make the initial guess that $y_{(x_1)}$ changes from $y_{0(x_1)}$ to

$$y_{0(x_1)} + v_{(x_1)}\kappa,$$

and similarly $y_{(x_2)}$ becomes

$$y_{0(x_2)} + v_{(x_2)}\kappa.$$

But that cannot be the whole story. The fact that $y_{(x_2)}$ is now different will influence $y_{(x_1)}$ through the equation $y_{(x_1)} = g_{(x_1)}(y_{(x_2)}, \kappa)$, and vice versa. So a better guess for $y_{(x_1)}$ and $y_{(x_2)}$ is

$$y_{0(x_1)} + v_{(x_1)}\kappa + G_{(x_1, x_2)}v_{(x_2)}\kappa$$

and

$$y_{0(x_2)} + v_{(x_2)}\kappa + G_{(x_2, x_1)}v_{(x_1)}\kappa.$$

Repeating this logic indefinitely, we would get a candidate expression for $y_{(x_1)}$ and $y_{(x_2)}$ in the form of an infinite series. It can be succinctly written as

$$y = y_0 + \kappa \sum_{n=0}^{\infty} G^n v + O(\kappa^2). \quad (\text{A.3})$$

This expression is of course correct, as can be seen by the standard method of com-

parative statics. Taking exact differentials of the equations of the problem, we obtain

$$dy_{(x_1)} = G_{(x_1, x_2)} dy_{(x_2)} + v_{(x_1)} d\kappa,$$

$$dy_{(x_2)} = G_{(x_2, x_1)} dy_{(x_1)} + v_{(x_2)} d\kappa,$$

and in matrix notation,

$$(1 - G) (y - y_0) = v\kappa + O(\kappa^2),$$

$$y = y_0 + \kappa (1 - G)^{-1} v + O(\kappa^2).$$

This is equivalent to the candidate expression above, thanks to the matrix geometric series identity $(1 - G)^{-1} = \sum_{n=0}^{\infty} G^n$.

We see that to succeed in this kind of task, one must be able to invert the matrix $1 - G$, or equivalently, to sum an infinite series of powers of G . In the two-variable case this is not a problem, of course. When the number of variables is large, however, this becomes a major obstacle.

There is a way to proceed, however. In situations where we can easily diagonalize G , computing $(1 - G)^{-1}$ is straightforward. Diagonalization of G means that we can express it as

$$G = C^{-1}DC,$$

where C is a known matrix and D is a known diagonal matrix with eigenvalues of G on its diagonal: $D = \text{diag}(d_1, d_2, \dots)$. In this case

$$(1 - G)^{-1} = C^{-1} (1 - D)^{-1} C, \quad (1 - D)^{-1} = \text{diag} \left(\frac{1}{1 - d_1}, \frac{1}{1 - d_2}, \dots \right).$$

This strategy is extensively used in the main text. The action of the matrix C can be thought of as a change of basis in the vector space of infinitesimal changes in endogenous variables. In concrete examples it corresponds to either Fourier series expansion, or to spherical harmonic expansion. In those cases, G, C , and D are infinite-dimensional. One could also consider the case of discrete Fourier transform where the matrices would be finite-dimensional, but for brevity that case will be omitted.

A.3 Derivation of equation (1.18)

Start with the GDP equation (1.17) with trade costs characterized by the function $(1 - \kappa b(x, x')) T(x, x')$. Using the Taylor expansion $y(x) = y_0(x) + \kappa y_1(x) + O(\kappa^2)$ on the right-hand side of the GDP equation yields

$$\begin{aligned} y^\sigma(x) &= y_0^\sigma(x) - \kappa \int \frac{b(x, x') T(x, x') y_0(x')}{\int T(x'', x') y_0^{1-\sigma}(x'') dL(x'')} dL(x') \\ &\quad + \kappa \int T(x, x') y_0(x') \frac{\int b(x'', x') T(x'', x') y_0^{1-\sigma}(x'') dL(x'')}{\left(\int T(x'', x') y_0^{1-\sigma}(x'') dL(x'') \right)^2} dL(x') \\ &\quad + \sigma \kappa y_0^{\sigma-1}(x) (G y_1)(x) + O(\kappa^2). \end{aligned}$$

Remembering the expression (1.12) for $G_c(x, x')$, this can be written as

$$\begin{aligned} y^\sigma(x) &= y_0^\sigma(x) - \kappa \sigma y_0^{\sigma-1}(x) \int b(x, x') G_c(x, x') y_0(x') dL(x') \\ &\quad + \kappa \sigma^2 y_0^{\sigma-1}(x) \int G_c(x, x') \left(\int b(x'', x') G_c(x'', x') dL(x'') \right) y_0(x') dL(x') \\ &\quad + \sigma y_0^{\sigma-1}(x) (G y_1)(x) + O(\kappa^2). \end{aligned}$$

Taylor expanding the left hand side then leads to the final equation,

$$y_1(x) = - \int b(x, x') G_c(x, x') y_0(x') dL(x') \\ + \sigma \int G_c(x, x') \left(\int b(x'', x') G_c(x'', x') dL(x'') \right) y_0(x') dL(x') + (Gy_1)(x).$$

Given the definition (1.19), this is equivalent to (1.18).

A.4 Local-price-index-adjusted GDP

The welfare of individual agents is characterized by the local GDP per capita adjusted for the local price index, $y^{(P)}(x) \equiv y(x) / P(x)$, with

$$P(x) \equiv \frac{\sigma - 1}{\sigma} \left(\frac{1}{\sigma F} \int (1 - \kappa b(x', x)) T(x', x) y^{1-\sigma}(x') dL(x') \right)^{\frac{1}{1-\sigma}}.$$

The first-order Taylor expansion of $y^{(P)}(x)$ is

$$y^{(P)}(x) = y_0^{(P)}(x) + \kappa y_0^{(P)}(x) \frac{y_1(x)}{y_0(x)} + \frac{\kappa}{\sigma - 1} y_0^{(P)}(x) \frac{N_1(x)}{N(x)} + O(\kappa^2),$$

where

$$\tilde{N}(x) \equiv \int (1 - \kappa b(x', x)) T(x', x) y^{1-\sigma}(x') dL(x'),$$

$$N(x) \equiv \int T(x', x) y_0^{1-\sigma}(x') dL(x'),$$

$$N_1(x) \equiv \lim_{\kappa \rightarrow 0} \frac{1}{\kappa} \frac{\tilde{N}(x) - N(x)}{N(x)}.$$

Given the definition $y_1^{(P)}(x) \equiv \lim_{\kappa \rightarrow 0} (y^{(P)}(x) - y_0^{(P)}(x)) / \kappa$, this implies

$$\frac{y_1^{(P)}(x)}{y_0^{(P)}(x)} = \frac{y_1(x)}{y_0(x)} + \frac{1}{\sigma - 1} \frac{N_1(x)}{N(x)},$$

$$\begin{aligned} \frac{y_1^{(P)}(x)}{y_0^{(P)}(x)} &= \frac{y_1(x)}{y_0(x)} - \frac{\int T(x', x) y_0^{-\sigma}(x') y_1(x') dL(x')}{\int T(x', x) y_0^{1-\sigma}(x') dL(x')} \\ &\quad - \frac{1}{\sigma-1} \frac{\int b(x', x) T(x', x) y_0^{1-\sigma}(x') dL(x')}{\int T(x', x) y_0^{1-\sigma}(x') dL(x')}. \end{aligned}$$

Using the expression (1.12) for $G_c(x', x)$, this is

$$\frac{y_1^{(P)}(x)}{y_0^{(P)}(x)} = \frac{y_1(x)}{y_0(x)} - \sigma \int G_c(x', x) \frac{y_1(x')}{y_0(x')} dL(x') - \frac{\sigma}{\sigma-1} \hat{g}_c(x),$$

or in operator notation

$$\frac{y_1^{(P)}}{y_0^{(P)}} = (1 - \sigma G_c) \frac{y_1}{y_0} - \frac{\sigma}{\sigma-1} \hat{g}_c.$$

The function \hat{g}_c is defined in (1.26). If $G_c(x', x) = G_c(x, x')$ and $b(x', x) = b(x, x')$, then $\hat{g}_c = \tilde{g}_c$, and we can write

$$\frac{y_1^{(P)}}{y_0^{(P)}} = (1 - \sigma G_c) \frac{y_1}{y_0} - \frac{\sigma}{\sigma-1} \tilde{g}_c.$$

A.5 Properties of $G_c^{*2}(\theta)$ on the circle for large R

In the present context, the expression (1.12) for G_c becomes

$$G_c(\theta) = \frac{1}{\sigma L \bar{T}} T(\theta), \tag{A.4}$$

with the average $T(\theta)$ defined as $\bar{T} \equiv L^{-1} \int_{-\pi}^{\pi} T(\theta) dL(\theta)$. The function of interest, $G_c^{*2}(\theta)$, can be rewritten as

$$\begin{aligned} G_c^{*2}(\theta) &\equiv \int_{-\pi}^{\pi} G_c(\theta') G_c(\theta - \theta') d\theta' = \frac{1}{(\sigma L \bar{T})^2} T^{*2}(\theta) \\ &\equiv \frac{1}{(\sigma L \bar{T})^2} \int_{-\pi}^{\pi} T(\theta') T(\theta - \theta') d\theta'. \end{aligned}$$

The transportation costs $\tilde{\tau}(d)$ are asymptotically power-law, in the sense of satisfying condition (1.29). This implies that $T(\theta)$ is asymptotically power-law as well,

$$a_h^{1-\sigma} \hat{T}(\theta) \leq T(\theta) \leq a_l^{1-\sigma} \hat{T}(\theta), \quad (\text{A.5})$$

with $\hat{T}(d) = \hat{\tau}^{1-\sigma}(d)$. Since

$$\int_{-\pi}^{\pi} \hat{T}(\theta) d\theta = 2 \int_0^{\frac{1}{\alpha R}} d\theta + 2 \int_{\frac{1}{\alpha R}}^{\pi} (\alpha R \theta)^{\rho(1-\sigma)} d\theta = \frac{2\delta}{2\delta-1} \frac{2}{\alpha R} - \frac{2\pi}{2\delta-1} \frac{1}{(\pi \alpha R)^{2\delta}},$$

a two sided bound on \bar{T} immediately follows,

$$\frac{a_h^{1-\sigma}}{2\delta-1} \left(\frac{2\delta}{\pi \alpha R} - \frac{1}{(\pi \alpha R)^{2\delta}} \right) \leq \bar{T} \leq \frac{a_l^{1-\sigma}}{2\delta-1} \left(\frac{2\delta}{\pi \alpha R} - \frac{1}{(\pi \alpha R)^{2\delta}} \right). \quad (\text{A.6})$$

Now that we know roughly the magnitude of \bar{T} , it remains to understand the nature of $T^{*2}(\theta)$. For this purpose, notice that (A.5) implies also

$$a_h^{1-\sigma} (T * \hat{T})(\theta) \leq T^{*2}(\theta) \leq a_l^{1-\sigma} (T * \hat{T})(\theta) \quad (\text{A.7})$$

and

$$a_h^{2-2\sigma} \hat{T}^{*2}(\theta) \leq T^{*2}(\theta) \leq a_l^{2-2\sigma} \hat{T}^{*2}(\theta). \quad (\text{A.8})$$

A.5.1 The peaks of $G_c^{*2}(\theta)$

Equation (A.4) implies

$$G_c^{*2}(0) = \frac{1}{(\sigma L \bar{T})^2} T^{*2}(0).$$

The related function $\hat{T}^2(\theta)$ can be integrated explicitly,

$$\hat{T}^{*2}(0) \equiv \int_{-\pi}^{\pi} \hat{T}^2(\theta) d\theta = \int_0^{\frac{1}{\alpha R}} d\theta + 2 \int_{\frac{1}{\alpha R}}^{\pi} \frac{1}{(\alpha R \theta)^{4\delta}} d\theta = \frac{4\delta}{4\delta-1} \frac{2}{\alpha R} + \frac{1}{1-4\delta} \frac{2\pi}{(\pi \alpha R)^{4\delta}}.$$

In the final expression, the first term comes from $|\theta|$ of order $1/(\alpha R)$. The part of the integration domain responsible for the second term is characterized by $|\theta|$ being of order one. The first term is important for $\delta \in (\frac{1}{4}, \infty)$. The second term dominates if $\delta \in (0, \frac{1}{4})$. Combining these two equalities with (A.8) and (A.6) gives

$$\begin{aligned} & \frac{(2\delta-1)^2 4\pi}{(4\delta-1)(\sigma L)^2} \frac{\frac{2\delta}{\pi \alpha R} - \frac{1}{(\pi \alpha R)^{4\delta}}}{\left(\frac{2\delta}{\pi \alpha R} - \frac{1}{(\pi \alpha R)^{2\delta}}\right)^2} \frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \\ & \leq G_c^{*2}(0) \leq \frac{(2\delta-1)^2 4\pi}{(4\delta-1)(\sigma L)^2} \frac{\frac{2\delta}{\pi \alpha R} - \frac{1}{(\pi \alpha R)^{4\delta}}}{\left(\frac{2\delta}{\pi \alpha R} - \frac{1}{(\pi \alpha R)^{2\delta}}\right)^2} \frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}. \end{aligned}$$

Alternatively, using also (A.4) and (A.5) with $\theta = 0$, the same relations imply

$$\frac{2\delta-1}{\sigma L} \frac{\frac{2\delta}{\pi \alpha R} - \frac{1}{(\pi \alpha R)^{4\delta}}}{\frac{2\delta}{\pi \alpha R} - \frac{1}{(\pi \alpha R)^{2\delta}}} G_c(0) \frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \leq G_c^{*2}(0) \leq \frac{2\delta-1}{\sigma L} \frac{\frac{2\delta}{\pi \alpha R} - \frac{1}{(\pi \alpha R)^{4\delta}}}{\frac{2\delta}{\pi \alpha R} - \frac{1}{(\pi \alpha R)^{2\delta}}} G_c(0) \frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}.$$

Specializing to various ranges for δ and remembering that R is large, the last two sets of inequalities imply (1.33) and (1.34).

A.5.2 $\delta < \frac{1}{2}$, tails of $G_c^{*2}(\theta)$

Consider θ greater than $2/(\alpha R)$. For simplicity, assume also that it is smaller than $\pi - 1/(\alpha R)$. Then the definition of $\hat{T}^{*2}(\theta)$ gives

$$\begin{aligned}\hat{T}^{*2}(\theta) = & \frac{1}{(\alpha R)^{4\delta}} \int_{-\pi}^{-\frac{1}{\alpha R}} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' + \frac{1}{(\alpha R)^{2\delta}} \int_{-\frac{1}{\alpha R}}^{\frac{1}{\alpha R}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' \\ & + \frac{1}{(\alpha R)^{4\delta}} \int_{\frac{1}{\alpha R}}^{\theta - \frac{1}{\alpha R}} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' + \frac{1}{(\alpha R)^{2\delta}} \int_{\theta - \frac{1}{\alpha R}}^{\theta + \frac{1}{\alpha R}} \frac{1}{|\theta'|^{2\delta}} d\theta' \\ & + \frac{1}{(\alpha R)^{4\delta}} \int_{\theta + \frac{1}{\alpha R}}^{\pi} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta'.\end{aligned}$$

It is easy to see that since R is large, the second and the fourth term give a contribution that is negligible relative to the remaining terms.

$$\begin{aligned}(\alpha R)^{4\delta} \hat{T}^{*2}(\theta) \approx & \int_{-\pi}^{-\frac{1}{\alpha R}} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' \\ & + \int_{\frac{1}{\alpha R}}^{\theta - \frac{1}{\alpha R}} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' + \int_{\theta + \frac{1}{\alpha R}}^{\pi} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' .\end{aligned}$$

Similarly, the remaining integrals will not change much if in their limits $1/(\alpha R)$ is replaced by zero. In that case, the three integrals can be merged into one.

$$(\alpha R)^{4\delta} \hat{T}^{*2}(\theta) \approx \int_{-\pi}^{\pi} \frac{1}{|\theta'|^{2\delta}} \frac{1}{|\theta - \theta'|^{2\delta}} d\theta' \equiv I(\theta). \quad (\text{A.9})$$

The integral¹ $I(\theta)$ is independent of R . The last relation, together with (A.4), (A.8), and (A.6), gives

$$\frac{(1-2\delta)^2}{(\sigma L)^2} \pi^{4\delta} I(\theta) \frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \lesssim G_c^{*2}(\theta) \lesssim \frac{(1-2\delta)^2}{(\sigma L)^2} \pi^{4\delta} I(\theta) \frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}.$$

This results, in turn, implies the first lines of (1.35) and (1.36).

$$A.5.3 \quad \delta > \frac{1}{2}, \text{ tails of } G_c^{*2}(\theta)$$

The definition of $(T * \hat{T})(\theta)$ is

$$(T * \hat{T})(\theta) = \int_{-\frac{1}{\alpha R}}^{\frac{1}{\alpha R}} T(\theta - \theta') d\theta' + \int_{\frac{1}{\alpha R}}^{\pi} \frac{T(\theta - \theta')}{|\alpha R \theta'|^{2\delta}} d\theta' + \int_{-\pi}^{-\frac{1}{\alpha R}} \frac{T(\theta - \theta')}{|\alpha R \theta'|^{2\delta}} d\theta'.$$

Assuming for simplicity that T is differentiable (this assumption can be lifted at the cost of a longer explanation), and integrating by parts, we get

$$\begin{aligned} (T * \hat{T})(\theta) &= \int_{-\frac{1}{\alpha R}}^{\frac{1}{\alpha R}} T(\theta - \theta') d\theta' \\ &+ \frac{1}{2\delta-1} \frac{1}{\alpha R} \left(T\left(\theta - \frac{1}{\alpha R}\right) + T\left(-\theta - \frac{1}{\alpha R}\right) \right) \\ &- \frac{1}{(\pi \alpha R)^{2\delta}} \frac{\pi}{2\delta-1} (T(\theta - \pi) + T(-\theta - \pi)) \\ &- \frac{1}{2\delta-1} \frac{1}{(\alpha R)^{2\delta}} \int_{\frac{1}{\alpha R}}^{\pi} |\theta'|^{1-2\delta} (T'(\theta - \theta') + T'(-\theta - \theta')) d\theta'. \end{aligned}$$

Consider θ in absolute value much greater than $1/(\alpha R)$. In that case, T is slowly varying. We can neglect the last two terms because they go to zero faster than $1/R$.

¹ The integral can be expressed using the gamma function Γ and the incomplete beta function B as $(-1)^{2\delta} |\theta|^{1-2\delta} \left(B_{\frac{\pi}{|\theta|}}(1-2\delta, 1-2\delta) - B_{-\frac{\pi}{|\theta|}}(1-2\delta, 1-2\delta) \right) + \left((-1)^{2\delta} - 1 \right) \sqrt{\pi} 2^{4\delta} \frac{\delta \Gamma(-2\delta)}{\Gamma(\frac{3}{2}-2\delta)} |\theta|^{1-4\delta}.$

In the remaining terms, we can approximate

$$\int_{-\frac{1}{\alpha R}}^{\frac{1}{\alpha R}} T(\theta - \theta') d\theta' \approx \frac{2}{\alpha R} T(\theta), \quad T\left(\theta - \frac{1}{\alpha R}\right) \approx T\left(-\theta - \frac{1}{\alpha R}\right) \approx T(\theta),$$

leading to the result that

$$(T * \hat{T})(\theta) \approx \frac{2\delta}{2\delta - 1} \frac{2}{\alpha R} T(\theta).$$

Inequality (A.7) then becomes

$$a_h^{1-\sigma} \frac{2\delta}{2\delta - 1} \frac{2}{\alpha R} T(\theta) \lesssim T^{*2}(\theta) \lesssim a_l^{1-\sigma} \frac{2\delta}{2\delta - 1} \frac{2}{\alpha R} T(\theta).$$

Using (A.4) and (A.6), the implication for $G_c^{*2}(\theta)$ is

$$\frac{1}{\sigma \rho_L} \left(\frac{a_l}{a_h}\right)^{\sigma-1} G_c(\theta) \lesssim G_c^{*2}(\theta) \lesssim \frac{1}{\sigma \rho_L} \left(\frac{a_h}{a_l}\right)^{\sigma-1} G_c(\theta). \quad (\text{A.10})$$

This implies the second line of (1.36). Combining this with (A.4), (A.6), and (A.5) then also gives the second line of (1.35).

A.6 Properties of $G_c^{*2}(\theta)$ on the sphere for large R

In analogy to the case of the circle, $G_c(\theta) = T(\theta) / (\sigma L \bar{T})$, where $T(\theta)$ averaged over the sphere is $\bar{T} \equiv L^{-1} \int_{\theta=0}^{\pi} \int_{\varphi=0}^{2\pi} T(\theta) dL(\theta, \varphi)$. The function $G_c^{*2}(\theta)$ is defined as

$$G_c^{*2}(\theta) \equiv \int_0^{\pi} \int_0^{2\pi} G_c(\theta') G_c(d(\theta, \varphi, \theta', \varphi')) \sin \theta' d\varphi' d\theta'.$$

Note that the right-hand side is independent of φ . It is again simple to find a bound on \bar{T} . An upper bound comes from (A.5), $\sin \theta \leq \theta$, and explicit integration.

$$\bar{T}a_l^{\sigma-1} \leq \frac{1}{2} \int_0^\pi \hat{T}(\theta) \sin \theta d\theta \leq \frac{1}{2} \int_0^\pi \hat{T}(\theta) \theta d\theta = \frac{1}{2} \int_0^{\frac{1}{\alpha R}} \theta d\theta + \frac{1}{2} \frac{1}{(\alpha R)^{2\delta}} \int_{\frac{1}{\alpha R}}^\pi \theta^{1-2\delta} d\theta.$$

$$\bar{T} \leq \left(\frac{1}{4} \frac{\delta}{\delta-1} \frac{1}{(\alpha R)^2} + \frac{1}{4} \frac{1}{1-\delta} \frac{\pi^2}{(\pi \alpha R)^{2\delta}} \right) a_l^{1-\sigma}.$$

A lower bound can be obtained by direct analogy, this time using $\sin \theta \geq 2/\pi \theta$, $\theta \in [0, \frac{\pi}{2}]$ rather than $\sin \theta \leq \theta$.

$$\begin{aligned} \bar{T}a_h^{\sigma-1} &\geq \frac{1}{2} \int_0^\pi \hat{T}(\theta) \sin \theta d\theta \\ &\geq \frac{1}{\pi} \int_0^{\frac{1}{\alpha R}} \theta d\theta + \frac{1}{\pi} \frac{1}{(\alpha R)^{2\delta}} \int_{\frac{1}{\alpha R}}^{\frac{\pi}{2}} \theta^{1-2\delta} d\theta + \frac{1}{2} \frac{1}{(\alpha R)^{2\delta}} \int_{\frac{\pi}{2}}^\pi \theta^{-2\delta} \sin \theta d\theta. \end{aligned}$$

Omitting the last term, which is positive, and evaluating the others yields

$$\bar{T}a_h^{\sigma-1} \geq \frac{1}{2\pi} \frac{1}{(\alpha R)^2} + \frac{1}{\pi} \frac{1}{(\alpha R)^{2\delta}} \frac{1}{2-2\delta} \left(\frac{\pi}{2} \right)^{2-2\delta} - \frac{1}{\pi} \frac{1}{(\alpha R)^{2\delta}} \frac{1}{2-2\delta} \left(\frac{1}{\alpha R} \right)^{2-2\delta},$$

$$\bar{T} \geq \left(\frac{\delta}{\delta-1} \frac{1}{2\pi (\alpha R)^2} + \frac{2^{2\delta-3}\pi}{1-\delta} \frac{1}{(\pi \alpha R)^{2\delta}} \right) a_h^{1-\sigma}.$$

The reader can certainly derive stricter bounds, but to understand the dependence on R , these two are sufficient.

$$\left(\frac{\delta}{\delta-1} \frac{1}{2\pi (\alpha R)^2} + \frac{2^{2\delta-3}\pi}{1-\delta} \frac{1}{(\pi \alpha R)^{2\delta}} \right) a_h^{1-\sigma} \tag{A.11}$$

$$\leq \bar{T} \leq \left(\frac{1}{4} \frac{\delta}{\delta-1} \frac{1}{(\alpha R)^2} + \frac{1}{4} \frac{1}{1-\delta} \frac{\pi^2}{(\pi \alpha R)^{2\delta}} \right) a_l^{1-\sigma}. \tag{A.12}$$

A.6.1 The peaks of $G_c^{*2}(\theta)$

$$G_c^{*2}(0) = \frac{1}{(\sigma L \bar{T})^2} \int_0^\pi \int_0^{2\pi} T^2(\theta) \sin \theta d\varphi d\theta = \frac{2\pi}{(\sigma L \bar{T})^2} \int_0^\pi T^2(\theta) \sin \theta d\theta.$$

The upper bound is

$$\begin{aligned} \int_0^\pi \hat{T}^2(\theta) \sin \theta d\theta &\leq \int_0^{\frac{1}{\alpha R}} \theta d\theta + \frac{1}{(\alpha R)^{4\delta}} \int_{\frac{1}{\alpha R}}^\pi \theta^{1-4\delta} d\theta, \\ \int_0^\pi \hat{T}^2(\theta) \sin \theta d\theta &\leq \frac{1}{2} \frac{1}{(\alpha R)^2} + \frac{1}{(\alpha R)^{4\delta}} \frac{1}{2-4\delta} \pi^{2-4\delta} - \frac{1}{(\alpha R)^{4\delta}} \frac{1}{2-4\delta} \frac{1}{(\alpha R)^{2-4\delta}}, \\ \int_0^\pi \hat{T}^2(\theta) \sin \theta d\theta &\leq \frac{\delta}{2\delta-1} \frac{1}{(\alpha R)^2} + \frac{\pi^2}{2} \frac{1}{1-2\delta} \frac{1}{(\pi \alpha R)^{4\delta}}. \end{aligned}$$

The lower bound is

$$\begin{aligned} \int_0^\pi \hat{T}^2(\theta) \sin \theta d\theta &\geq \frac{2}{\pi} \int_0^{\frac{1}{\alpha R}} \theta d\theta + \frac{1}{(\alpha R)^{4\delta}} \frac{2}{\pi} \int_{\frac{1}{\alpha R}}^{\frac{\pi}{2}} \theta^{1-4\delta} d\theta, \\ \int_0^\pi \hat{T}^2(\theta) \sin \theta d\theta &\geq \frac{1}{\pi} \frac{1}{(\alpha R)^2} + \frac{1}{(\alpha R)^{4\delta}} \frac{1}{\pi} \frac{1}{1-2\delta} \left(\frac{\pi}{2}\right)^{2-4\delta} - \frac{1}{(\alpha R)^2} \frac{1}{1-2\delta} \frac{1}{\pi}, \\ \int_0^\pi \hat{T}^2(\theta) \sin \theta d\theta &\geq \frac{1}{\pi} \frac{2\delta}{2\delta-1} \frac{1}{(\alpha R)^2} + \frac{2^{4\delta-2} \pi^{1-4\delta}}{1-2\delta} \frac{1}{(\alpha R)^{4\delta}}. \end{aligned}$$

The bounds combined:

$$\begin{aligned} &\left(\frac{1}{\pi} \frac{2\delta}{2\delta-1} \frac{1}{(\alpha R)^2} + \frac{2^{4\delta-2} \pi^{1-4\delta}}{1-2\delta} \frac{1}{(\alpha R)^{4\delta}} \right) a_h^{2-2\sigma} \\ &\leq \int_0^\pi T^2(\theta) \sin \theta d\theta \leq \left(\frac{\delta}{2\delta-1} \frac{1}{(\alpha R)^2} + \frac{\pi^2}{2} \frac{1}{1-2\delta} \frac{1}{(\pi \alpha R)^{4\delta}} \right) a_l^{2-2\sigma}. \end{aligned}$$

Combining this with (A.11) gives

$$\begin{aligned} & \frac{2\pi}{(\sigma L)^2} \frac{\frac{1}{\pi} \frac{2\delta}{2\delta-1} \frac{1}{(\alpha R)^2} + \frac{2^{4\delta-2} \pi^{1-4\delta}}{1-2\delta} \frac{1}{(\alpha R)^{4\delta}}}{\left(\frac{1}{4} \frac{\delta}{\delta-1} \frac{1}{(\alpha R)^2} + \frac{1}{4} \frac{1}{1-\delta} \frac{\pi^2}{(\pi \alpha R)^{2\delta}} \right)^2} \frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \\ & \leq G_c^{*2}(0) \leq \frac{2\pi}{(\sigma L)^2} \frac{\frac{\delta}{2\delta-1} \frac{1}{(\alpha R)^2} + \frac{\pi^2}{2} \frac{1}{1-2\delta} \frac{1}{(\pi \alpha R)^{4\delta}}}{\left(\frac{\delta}{\delta-1} \frac{1}{2\pi(\alpha R)^2} + \frac{2^{2\delta-3}\pi}{1-\delta} \frac{1}{(\pi \alpha R)^{2\delta}} \right)^2} \frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}, \end{aligned}$$

or alternatively, also with (A.5)

$$\begin{aligned} & \frac{2\pi G_c(0)}{\sigma L} \frac{\frac{1}{\pi} \frac{2\delta}{2\delta-1} \frac{1}{(\alpha R)^2} + \frac{2^{4\delta-2} \pi^{1-4\delta}}{1-2\delta} \frac{1}{(\alpha R)^{4\delta}}}{\frac{1}{4} \frac{\delta}{\delta-1} \frac{1}{(\alpha R)^2} + \frac{1}{4} \frac{1}{1-\delta} \frac{\pi^2}{(\pi \alpha R)^{2\delta}}} \frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \\ & \leq G_c^{*2}(0) \leq \frac{2\pi G_c(0)}{\sigma L} \frac{\frac{\delta}{2\delta-1} \frac{1}{(\alpha R)^2} + \frac{\pi^2}{2} \frac{1}{1-2\delta} \frac{1}{(\pi \alpha R)^{4\delta}}}{\frac{\delta}{\delta-1} \frac{1}{2\pi(\alpha R)^2} + \frac{2^{2\delta-3}\pi}{1-\delta} \frac{1}{(\pi \alpha R)^{2\delta}}} \frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}. \end{aligned}$$

A.6.2 $\delta < 1$, tails of $G_c^{*2}(\theta)$

$$\hat{T}^2(\theta) \equiv \int_0^\pi \int_0^{2\pi} \hat{T}(\theta') \hat{T}(d(\theta, \varphi, \theta', \varphi')) \sin \theta' d\varphi' d\theta'.$$

This integral contains regions where either θ' or $d(\theta, \varphi, \theta', \varphi')$ are smaller than $1/(\alpha R)$.

As in the case of the circle with $\delta < \frac{1}{2}$, they do not contribute much to the integral when $d(\theta, \varphi, \theta', \varphi') \gg 1/(\alpha R)$, and in this case can be safely ignored. As a result, we get the following approximation.

$$(\alpha R)^{4\delta} \hat{T}^2(\theta) \approx \int_0^\pi \int_0^{2\pi} \frac{1}{\theta'^{2\delta}} \frac{1}{d^{2\delta}(\theta, \varphi, \theta', \varphi')} \sin \theta' d\varphi' d\theta' \equiv I_{(2)}(\theta).$$

The right-hand side is now independent of R . Together with (A.11) this implies

$$\frac{2^5 \pi^{4\delta-3}}{(\sigma L)^2} (1-\delta) I_{(2)}(\theta) \frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \leq G_c^{*2}(\theta) \leq \frac{2^5 \pi^{4\delta-3}}{(\sigma L)^2} (1-\delta) I_{(2)}(\theta) \frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}.$$

A.6.3 $\delta > 1$, tails of $G_c^{*2}(\theta)$

$$(T * \hat{T})(\theta) \equiv \int_0^\pi \int_0^{2\pi} \hat{T}(\theta') T(d(\theta, \varphi, \theta', \varphi')) \sin \theta' d\varphi' d\theta',$$

$$\begin{aligned} (T * \hat{T})(\theta) &= \int_0^{\frac{1}{\alpha R}} \int_0^{2\pi} T(d(\theta, \varphi, \theta', \varphi')) \sin \theta' d\varphi' d\theta' \\ &\quad + \frac{1}{(\alpha R)^{2\delta}} \int_{\frac{1}{\alpha R}}^\pi \int_0^{2\pi} \frac{1}{\theta'^{2\delta}} T(d(\theta, \varphi, \theta', \varphi')) \sin \theta' d\varphi' d\theta'. \end{aligned}$$

For $\theta \gg 1/(\alpha R)$, T is slowly varying.

$$\begin{aligned} (T * \hat{T})(\theta) &\approx \frac{\pi}{(\alpha R)^2} T(\theta) \\ &\quad + \frac{1}{(\alpha R)^{2\delta}} \int_{\frac{1}{\alpha R}}^\pi \int_0^{2\pi} \frac{1}{\theta'^{2\delta}} T(d(\theta, \varphi, \theta', \varphi')) \sin \theta' d\varphi' d\theta'. \end{aligned}$$

Using $\sin \theta < \theta$ and integrating by parts,

$$\begin{aligned} (T * \hat{T})(\theta) &\leq \frac{\pi}{(\alpha R)^2} T(\theta) \\ &\quad + \frac{1}{(\alpha R)^{2\delta}} \frac{1}{2-2\delta} \int_0^{2\pi} \left[\theta'^{2-2\delta} T(d(\theta, \varphi, \theta', \varphi')) \right]_{\frac{1}{\alpha R}}^\pi d\varphi' \\ &\quad - \frac{1}{(\alpha R)^{2\delta}} \frac{1}{2-2\delta} \int_{\frac{1}{\alpha R}}^\pi \int_0^{2\pi} \theta'^{2-2\delta} \frac{\partial T(d(\theta, \varphi, \theta', \varphi'))}{\partial \theta'} d\varphi' d\theta'. \end{aligned}$$

Since $\theta'^{2-2\delta}$ is small for $\theta' \gg 1/(\alpha R)$ and T varies slowly over the region where θ' is of order $1/(\alpha R)$, the last term can be neglected.

$$\begin{aligned} (T * \hat{T})(\theta) &\leq \frac{\pi}{(\alpha R)^2} T(\theta) \\ &\quad + \frac{1}{(\alpha R)^{2\delta}} \frac{1}{1-\delta} \pi^{3-2\delta} T(\pi - \theta) \\ &\quad - \frac{1}{(\alpha R)^{2\delta}} \frac{1}{2-2\delta} \int_0^{2\pi} \frac{1}{(\alpha R)^{1-2\delta}} T\left(d\left(\theta, \varphi, \frac{1}{\alpha R}, \varphi'\right)\right) \frac{1}{\alpha R} d\varphi'. \end{aligned}$$

In the last term, the dependence on φ' is very weak since θ' is set to $1/(\alpha R)$. We can approximate

$$\begin{aligned} (T * \hat{T})(\theta) &\leq \frac{\pi}{(\alpha R)^2} T(\theta) \\ &\quad + \frac{1}{(\alpha R)^{2\delta}} \frac{1}{1-\delta} \pi^{3-2\delta} T(\pi - \theta) \\ &\quad - \frac{1}{(\alpha R)^2} \frac{1}{1-\delta} \pi T(\theta). \end{aligned}$$

Since R is large

$$(T * \hat{T})(\theta) \leq \frac{\pi}{(\alpha R)^2} \frac{\delta}{\delta - 1} T(\theta).$$

A lower bound can be obtained analogously using $\sin \theta < \frac{2}{\pi} \theta$.

$$(T * \hat{T})(\theta) \geq \frac{\pi}{(\alpha R)^2} T(\theta) + \frac{1}{(\alpha R)^{2\delta}} \frac{1}{2-2\delta} \frac{2}{\pi} \int_0^{2\pi} \left[\theta'^{2-2\delta} T(d(\theta, \varphi, \theta', \varphi')) \right]_{\frac{1}{\alpha R}}^{\frac{\pi}{2}} d\varphi',$$

$$(T * \hat{T})(\theta) \geq \left(\pi + \frac{2}{\delta - 1} \right) \frac{1}{(\alpha R)^2} T(\theta).$$

Together with (A.11), this leads to

$$\frac{2}{\sigma L} \frac{(\pi(\delta-1)+2)}{\pi\delta} G_c(\theta) \frac{a_l^{\sigma-1}}{a_h^{\sigma-1}} \leq G_c^{*2}(\theta) \leq \frac{2}{\sigma L} G_c(\theta) \frac{a_h^{\sigma-1}}{a_l^{\sigma-1}},$$

or alternatively, to

$$\frac{8}{(\sigma L)^2} \frac{\delta-1}{\delta} \frac{(\pi(\delta-1)+2)}{\pi\delta} \frac{1}{(\pi\alpha R)^{2\delta-2}} \frac{a_l^{2\sigma-2}}{a_h^{2\sigma-2}} \leq G_c^{*2}(\pi) \leq \frac{4\pi}{(\sigma L)^2} \frac{\delta-1}{\delta} \frac{1}{(\pi\alpha R)^{2\delta-2}} \frac{a_h^{2\sigma-2}}{a_l^{2\sigma-2}}.$$

A.7 Fourier series expansions

A.7.1 Fourier series expansions of country indicator functions

In the case of the indicator function $1_{C_A}(\theta)$ of the set $C_A = (-\frac{\pi}{2}, \frac{\pi}{2})$ characterizing country A , the standard formula for Fourier coefficients (1.38) specializes to

$$1_{C_A,n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} 1_{C_A}(\theta) e^{-in\theta} d\theta = \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} e^{-in\theta} d\theta.$$

Evaluating the integral in various cases,

$$1_{C_A,n} = \begin{cases} \frac{1}{2} & \text{for } n = 0, \\ 0 & \text{for } n \text{ even and nonzero,} \\ \frac{(-1)^{\frac{n-1}{2}}}{\pi n} & \text{for } n \text{ odd.} \end{cases} \quad (\text{A.13})$$

Now consider the indicator function $1_{C_B}(\theta)$ of country B with $C_B = (-\pi, -\frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi]$. Because almost everywhere $1_{C_A}(\theta) + 1_{C_B}(\theta) = 1$, it must be that $1_{C_A,0} +$

$1_{C_A,0} = 1$ and for nonzero n , $1_{C_A,n} + 1_{C_A,-n} = 0$. This implies

$$1_{C_B,n} = \begin{cases} \frac{1}{2} & \text{for } n = 0, \\ 0 & \text{for } n \text{ even and nonzero,} \\ \frac{(-1)^{\frac{n+1}{2}}}{\pi n} & \text{for } n \text{ odd.} \end{cases} \quad (\text{A.14})$$

We see that the Fourier series expansions of the country indicator functions are

$$\begin{aligned} 1_{C_A}(\theta) &= \frac{1}{2} + \frac{1}{\pi} \sum_{n \in \mathbb{Z}, n \text{ odd}} (-1)^{\frac{n-1}{2}} \frac{1}{n} e^{in\theta}, \\ 1_{C_B}(\theta) &= \frac{1}{2} - \frac{1}{\pi} \sum_{n \in \mathbb{Z}, n \text{ odd}} (-1)^{\frac{n-1}{2}} \frac{1}{n} e^{in\theta}. \end{aligned} \quad (\text{A.15})$$

For future convenience, multiply the expression for $1_{C_A}(\theta)$ by $e^{im\theta}$ and replace $n \rightarrow n - m$ to arrive at the following identity

$$1_{C_A}(\theta) e^{im\theta} = \frac{1}{2} e^{im\theta} + \frac{1}{\pi} \sum_{n \in \mathbb{Z}, n-m \text{ odd}} (-1)^{\frac{n-m+1}{2}} \frac{1}{n-m} e^{in\theta}. \quad (\text{A.16})$$

A.7.2 Fourier expansion of $\tilde{h}(\theta)$

For a symmetric function $H(\theta)$ on the circle (extended periodically to the real line), let us evaluate $\tilde{h}_n \equiv (\tilde{H}1)_n$ where \tilde{H} is the integral operator with kernel $\rho_L \tilde{H}(d(\theta, \theta')) b(\theta, \theta')$. The function $b(\theta, \theta')$ is one whenever θ and θ' lie on opposite sides of the border and zero otherwise. It can be written as $b(\theta, \theta') = b_{AB}(\theta, \theta') + b_{BA}(\theta, \theta')$ with $b_{AB}(\theta, \theta') \equiv 1_{C_A}(\theta) 1_{C_B}(\theta')$ and $b_{BA}(\theta, \theta') = 1_{C_B}(\theta) 1_{C_A}(\theta')$. For ease of notation, define also $\tilde{h}_{AB} \equiv \tilde{H}_{AB}1$ with the kernel of the operator \tilde{H}_{AB} being $\rho_L \tilde{H}(d(\theta, \theta')) b_{AB}(\theta, \theta')$, and analogously $\tilde{h}_{BA} \equiv \tilde{H}_{BA}1$. These two functions add up to \tilde{h} , so $\tilde{h}_n = \tilde{h}_{AB,n} + \tilde{h}_{BA,n}$.

First, compute $\tilde{h}_{AB,n}$.

$$\tilde{h}_{AB}(\theta) = \rho_L 1_{C_A}(\theta) \int_{-\pi}^{\pi} H(\theta - \theta') 1_{C_B}(\theta') d\theta'.$$

Fourier expanding the function $H(\theta - \theta')$,

$$\tilde{h}_{AB}(\theta) = \rho_L 1_{C_A}(\theta) \sum_{m \in \mathbb{Z}} H_m e^{im\theta} \int_{-\pi}^{\pi} e^{-im\theta'} 1_{C_B}(\theta') d\theta' = L \sum_{m \in \mathbb{Z}} H_m 1_{C_B,m} 1_{C_A}(\theta) e^{im\theta}.$$

Note that because $H(\theta)$ is symmetric, $H_m = H_{-m}$. Substituting for $1_{C_A}(\theta) e^{im\theta}$ from (A.16) gives

$$\tilde{h}_{AB}(\theta) = L \sum_{m \in \mathbb{Z}} A_m 1_{C_B,m} \left(\frac{1}{2} e^{im\theta} + \frac{1}{\pi} \sum_{n \in \mathbb{Z}, n-m \text{ odd}} (-1)^{\frac{n-m+1}{2}} \frac{1}{n-m} e^{in\theta} \right).$$

Exchanging the order of summations,

$$\begin{aligned} \tilde{h}_{AB}(\theta) &= \frac{1}{2} L \sum_{m \in \mathbb{Z}} H_m 1_{C_B,m} e^{im\theta} \\ &+ \frac{1}{\pi} L \sum_{n \in \mathbb{Z}} \left(\sum_{m \in \mathbb{Z}, m-n \text{ odd}} (-1)^{\frac{n-m+1}{2}} \frac{1}{n-m} H_m 1_{C_B,m} \right) e^{in\theta}. \end{aligned}$$

The Fourier series expansion $\tilde{h}_{BA}(\theta)$ follows from the one for $\tilde{h}_{AB}(\theta)$ because these two functions are related to each other by the shift $\theta \rightarrow \theta + \pi$,

$$\begin{aligned} \tilde{h}_{BA}(\theta) &= \frac{1}{2} L \sum_{m \in \mathbb{Z}} (-1)^m H_m 1_{C_B,m} e^{im\theta} \\ &+ \frac{1}{\pi} L \sum_{n \in \mathbb{Z}} \left(\sum_{m \in \mathbb{Z}, m-n \text{ odd}} (-1)^{\frac{-n-m+1}{2}} \frac{1}{n-m} H_m 1_{C_B,m} \right) e^{in\theta}. \end{aligned}$$

According to (A.14), $1_{C_B,m}$ with even m is nonzero only for $m = 0$, in which case

$1_{C_B,0} = \frac{1}{2}$. This means that after adding the two equations, we obtain

$$\begin{aligned}\tilde{h}(\theta) &= \frac{1}{2}LH_0 \\ &+ \frac{1}{\pi}L \sum_{n \in \mathbb{Z}} \left(\sum_{m \in \mathbb{Z}, m-n \text{ odd}} \frac{(-1)^{\frac{n-m+1}{2}} + (-1)^{\frac{-n-m+1}{2}}}{n-m} H_m 1_{C_B,m} \right) e^{in\theta}.\end{aligned}$$

This is the desired Fourier expansion of $\tilde{h}(\theta)$. From here we can read off the individual Fourier coefficients.

$$\tilde{h}_n = \begin{cases} \frac{1}{2}LH_0\delta_{0n} + \frac{2}{\pi}L \sum_{m \in \mathbb{Z}, m \text{ odd}} \frac{(-1)^{\frac{n-m+1}{2}}}{n-m} H_m 1_{C_B,m} & \text{for } n \text{ even,} \\ 0 & \text{for } n \text{ odd.} \end{cases}$$

Here δ_{0n} is the Kronecker delta, equal to one if $n = 0$ and zero otherwise. Now we can substitute the explicit expressions (A.14) for $1_{C_B,m}$ and use the relabeling

$$\begin{aligned}\sum_{m \in \mathbb{Z}, m \text{ odd}} \frac{1}{n-m} \frac{H_m}{m} &= \sum_{m \in \mathbb{Z}, m \text{ odd positive}} \left(\frac{1}{n-m} \frac{H_m}{m} - \frac{1}{n+m} \frac{H_m}{m} \right) \\ &= -2 \sum_{m=0}^{\infty} \frac{H_{2m+1}}{(2m+1)^2 - n^2}\end{aligned}$$

to get the final expression

$$\tilde{h}_n = \begin{cases} 0 & \text{for } n \text{ odd,} \\ \frac{1}{2}LH_0\delta_{0n} - \frac{4}{\pi^2}(-1)^{\frac{n}{2}} L \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2 - n^2} H_{2m+1} & \text{for } n \text{ even.} \end{cases}$$

A.7.3 Fourier expansion of $\tilde{g}_c(\theta)$

The discussion above was for an unspecified function $H(\theta)$ on the circle. Specializing to $G_c(\theta)$, we get the result

$$\tilde{g}_{c,n} = \begin{cases} 0 & \text{for } n \text{ odd,} \\ \frac{1}{2\sigma} \delta_{0n} - \frac{4}{\pi^2} (-1)^{\frac{n}{2}} L \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2 - n^2} G_{c,2m+1} & \text{for } n \text{ even.} \end{cases} \quad (\text{A.17})$$

A.8 Derivation of the expression for T_n on circle

The goal here is to evaluate the Fourier coefficients of

$$T(\theta) = \left(\frac{1}{1 + 4\alpha^2 R^2 \sin^2 \frac{\theta}{2}} \right)^\delta.$$

The standard formula (1.38) for Fourier coefficients implies

$$T_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{-in\theta}}{(1 + 4\alpha^2 R^2 \sin^2 \frac{\theta}{2})^\delta} d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\cos n\theta}{(1 + 4\alpha^2 R^2 \sin^2 \frac{\theta}{2})^\delta} d\theta,$$

where the second equality follows from the Euler formula $e^{ix} = \cos x + i \sin x$ and from the fact that $T(\theta)$ is symmetric while $\sin n\theta$ is antisymmetric. Taking advantage of the symmetry of the final integrand to adjust the integration range and using the identity $\sin^2(\theta/2) = (1 - \cos \theta)/2$,

$$T_n = \frac{1}{\pi} \int_0^\pi \frac{\cos n\theta}{(1 + 2\alpha^2 R^2 - 2\alpha^2 R^2 \cos \theta)^\delta} d\theta.$$

Define $Z \equiv 1/\sqrt{1+4\alpha^2 R^2}$. Then $2\alpha^2 R^2 = (1 - Z^2) / (2Z^2)$, and the integral can be rewritten as

$$T_n = Z^\delta \frac{1}{\pi} \left(\frac{2Z}{1+Z^2} \right)^\delta \int_0^\pi \frac{\cos n\theta}{\left(1 - \frac{1-Z^2}{1+Z^2} \cos \theta\right)^\delta} d\theta.$$

The (corrected² version) of second equation in paragraph 9.131 on p. 1008 of Gradshteyn and Ryzhik (2007) states that

$$P_\nu^m(z) = \frac{\nu(\nu-1)\dots(\nu-m+1)}{\pi} \int_0^\pi \frac{\cos m\varphi}{\left(z - \sqrt{z^2-1} \cos \varphi\right)^{\nu+1}} d\varphi,$$

where $P_\nu^m(z)$ denotes³ associated Legendre functions of the first kind. Using this equation with the replacement $\{m, \nu, \varphi\} \rightarrow \{n, \delta-1, \theta\}$, gives

$$\frac{(-1)^n}{(1-\delta)_n} P_{\delta-1}^n(z) = \frac{1}{\pi} \frac{1}{z^\delta} \int_0^\pi \frac{\cos n\theta}{\left(1 - \frac{\sqrt{z^2-1}}{z} \cos \theta\right)^\delta} d\theta,$$

where $(1-\delta)_n$ is the Pochhammer symbol. Replacing also $z \rightarrow \frac{1+Z^2}{2Z}$ and noticing that this corresponds to $\frac{\sqrt{z^2-1}}{z} \rightarrow \frac{1-Z^2}{1+Z^2}$, one gets the identity

$$\frac{(-1)^n}{(1-\delta)_n} P_{\delta-1}^n\left(\frac{1+Z^2}{2Z}\right) = \frac{1}{\pi} \left(\frac{2Z}{1+Z^2}\right)^\delta \int_0^\pi \frac{\cos n\theta}{\left(1 - \frac{1-Z^2}{1+Z^2} \cos \theta\right)^\delta} d\theta.$$

The integral on the right-hand side has the same form as the one in the expression for T_n , which leads to the conclusion

$$T_n = \frac{(-1)^n}{(1-\delta)_n} Z^\delta P_{\delta-1}^n\left(\frac{1+Z^2}{2Z}\right).$$

² The formula in the book contains an additional factor of $(-1)^m$, which is a typo.

³ The Mathematica notation for this function is LegendreP[$\nu, \mu, 3, z$].

A.9 Spherical harmonic expansions

A.9.1 Spherical harmonic expansions of country indicator functions

Let us find the spherical harmonic expansion of the indicator function $1_{C_A}(\theta)$ of the set $C_A = \{(\theta, \varphi) \mid \theta \in [0, \frac{\pi}{2})\}$, which corresponds to country A . The general formula for spherical harmonic coefficients (1.51) gives

$$(1_{C_A})_l^m = \int_0^{\frac{\pi}{2}} \int_0^{2\pi} Y_l^{m*}(\theta, \varphi) d\varphi \sin \theta d\theta.$$

This vanishes for non-zero m . For $m = 0$, we can use the expression (1.52) to simplify the integral to

$$(1_{C_A})_l^0 = \sqrt{\pi} \sqrt{2l+1} \int_0^{\frac{\pi}{2}} P_l(\cos \theta) \sin \theta d\theta = \sqrt{\pi} \sqrt{2l+1} \int_0^1 P_l(t) dt.$$

The last integral can be evaluated explicitly, with the result

$$(1_{C_A})_l^0 = \begin{cases} \sqrt{\pi} & \text{for } l = 0, \\ 0 & \text{for } l \text{ even and nonzero,} \\ \sqrt{\pi} \sqrt{2l+1} \frac{(-1)^{\frac{l-1}{2}}}{2^l} \frac{(l-1)!}{\frac{l-1}{2}! \frac{l+1}{2}!} & \text{for } l \text{ odd.} \end{cases}$$

Because up to a set of measure zero $1_{C_A}(\theta, \varphi) + 1_{C_B}(\theta, \varphi) = 1 = 2\sqrt{\pi}Y_0^0(\theta, \varphi)$, the spherical harmonic coefficients of 1_{C_B} with $C_B = \{(\theta, \varphi) \mid \theta \in (\frac{\pi}{2}, \pi]\}$ follow. $(1_{C_B})_l^m$ with non-zero m vanishes, and

$$(1_{C_B})_l^0 = \begin{cases} \sqrt{\pi} & \text{for } l = 0, \\ 0 & \text{for } l \text{ even and nonzero,} \\ \sqrt{\pi} \sqrt{2l+1} \frac{(-1)^{\frac{-l-1}{2}}}{2^l} \frac{(l-1)!}{\frac{l-1}{2}! \frac{l+1}{2}!} & \text{for } l \text{ odd.} \end{cases} \quad (\text{A.18})$$

A.9.2 Spherical harmonic expansion of $Y_{l'}^0(\theta, \varphi) C_B(\theta, \varphi)$ and

$$Y_{l'}^0(\theta, \varphi) C_A(\theta, \varphi)$$

To find spherical harmonic coefficients of $Y_{l'}^0(\theta, \varphi) 1_{C_B}(\theta, \varphi)$ we may again use (1.51). The coefficients with nonzero m vanish, because $Y_0^0(\theta, \varphi) 1_{C_B}(\theta, \varphi)$ is independent of φ . For the remaining coefficients,

$$[Y_{l'}^0 1_{C_B}]_l^0 = 2\pi \int_0^\pi Y_{l'}^0(\theta, 0) 1_{C_B}(\theta, 0) Y_l^0(\theta, 0) \sin \theta d\theta.$$

Due to (1.52) this is

$$[Y_{l'}^0 1_{C_B}]_l^0 = \frac{1}{2} \sqrt{2l'+1} \sqrt{2l+1} \int_0^1 P_{l'}(t) P_l(t) dt.$$

It is not hard to evaluate the integral for any given pair l, l' using the standard definition of Legendre polynomials. An alternative expression may be obtained as follows.

$$[Y_{l'}^0 1_{C_B}]_l^0 = 2\pi \sum_{l''=0}^{\infty} (1_{C_B})_{l''}^0 \int_0^\pi Y_{l'}^0(\theta, 0) Y_{l''}^0(\theta, 0) Y_l^0(\theta, 0) \sin \theta d\theta.$$

$$\begin{aligned} [Y_{l'}^0 1_{C_B}]_l^0 &= \frac{\sqrt{2l+1} \sqrt{2l'+1}}{\sqrt{4\pi}} \\ &\times \sum_{l''=0}^{\infty} (1_{C_B})_{l''}^0 \sqrt{2l''+1} \frac{1}{2} \int_0^\pi P_{l'}(\cos \theta) P_{l''}(\cos \theta) P_l(\cos \theta) \sin \theta d\theta. \end{aligned}$$

$$[Y_{l'}^0 1_{C_B}]_l^0 = \frac{\sqrt{2l+1}\sqrt{2l'+1}}{\sqrt{4\pi}} \sum_{l''=0}^{\infty} (1_{C_B})_{l''}^0 \sqrt{2l''+1} \begin{pmatrix} l' & l'' & l \\ 0 & 0 & 0 \end{pmatrix}^2,$$

where $\begin{pmatrix} l' & l'' & l \\ 0 & 0 & 0 \end{pmatrix}$ is the Wigner 3j symbol (closely related to Clebsch–Gordan coefficients). Substituting the explicit expressions (A.18) for $(1_{C_B})_{l''}^0$ leads to

$$\begin{aligned} [Y_{l'}^0 1_{C_B}]_l^0 &= \frac{\sqrt{2l+1}\sqrt{2l'+1}}{2} \begin{pmatrix} l' & 0 & l \\ 0 & 0 & 0 \end{pmatrix}^2 \\ &+ \frac{\sqrt{2l+1}\sqrt{2l'+1}}{2} \sum_{l''=1, l'' \text{ odd}}^{\infty} \frac{(-1)^{\frac{l''-1}{2}} (2l''+1) (l''-1)!}{2^{l''}} \frac{l''-1!}{2!} \frac{l''+1!}{2!} \begin{pmatrix} l' & l'' & l \\ 0 & 0 & 0 \end{pmatrix}^2. \end{aligned}$$

Because the Wigner 3j symbol vanishes whenever the triangle inequality between l, l' , and l'' is not satisfied, the infinite sum reduces to a finite one:

$$\begin{aligned} [Y_{l'}^0 1_{C_B}]_l^0 &= \frac{1}{2} \delta_{ll'} \\ &+ \frac{\sqrt{2l+1}\sqrt{2l'+1}}{2} \sum_{l''=|l+l'|, l'' \text{ odd}}^{l+l'} \frac{(-1)^{\frac{l''-1}{2}} (2l''+1) (l''-1)!}{2^{l''}} \frac{l''-1!}{2!} \frac{l''+1!}{2!} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}^2. \end{aligned}$$

Here $\delta_{ll'}$ is the Kronecker delta, equal to one when $l = l'$, and zero otherwise. Since up to a set of measure zero $1_{C_A}(\theta, \varphi) + 1_{C_B}(\theta, \varphi) = 1 = 2\sqrt{\pi}Y_0^0(\theta, \varphi)$, this result also implies

$$\begin{aligned} [Y_{l'}^0 1_{C_A}]_l^0 &= \frac{1}{2} \delta_{ll'} \\ &+ \frac{\sqrt{2l+1}\sqrt{2l'+1}}{2} \sum_{l''=|l+l'|, l'' \text{ odd}}^{l+l'} \frac{(-1)^{\frac{l''+1}{2}} (2l''+1) (l''-1)!}{2^{l''}} \frac{l''-1!}{2!} \frac{l''+1!}{2!} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}^2. \end{aligned}$$

A.9.3 Spherical harmonic expansions used to analyze the impact of border costs

Let us find certain spherical harmonic expansions needed to evaluate the impact of changes in border costs. The 'border indicator function' $b(x, x') \equiv b(\theta, \varphi, \theta', \varphi')$ may be decomposed into two parts

$$\begin{aligned} b(\theta, \varphi, \theta', \varphi') &= b_{AB}(\theta, \varphi, \theta', \varphi') + b_{BA}(\theta, \varphi, \theta', \varphi'), \\ b_{AB}(\theta, \varphi, \theta', \varphi') &\equiv 1_{C_A}(\theta, \varphi) 1_{C_B}(\theta', \varphi'), \\ b_{BA}(\theta, \varphi, \theta', \varphi') &\equiv 1_{C_B}(\theta, \varphi) 1_{C_A}(\theta', \varphi'). \end{aligned}$$

Consider a function $A(\theta, \varphi) \equiv A(\theta)$ on the sphere that is independent of φ . Denote the spherical angle (i.e. $1/R$ times the spherical distance) between points $x \equiv (\theta, \varphi)$ and $x' \equiv (\theta', \varphi')$ as $\tilde{d}(x, x') \equiv \tilde{d}(\theta, \varphi, \theta', \varphi')$. This angle can be computed with the help of the identity

$$\cos \tilde{d}(x, x') = \sin \theta \sin \theta' + \cos \theta \cos \theta' \sin(\varphi - \varphi').$$

The function whose spherical harmonic expansion we need to evaluate is $a(\theta, \varphi) \equiv a(\theta)$ defined⁴ by the equation

$$a(\theta) \equiv \frac{1}{L} \int A(\tilde{d}(\theta, \varphi, \theta', \varphi')) b(\theta, \varphi, \theta', \varphi') dL(\theta', \varphi').$$

⁴ Note that the integral on the right-hand side is independent of φ due to the rotational symmetry of each factor inside the integral.

It will be convenient to introduce also notation for its two parts corresponding to the decomposition of b in terms of b_{AB} and b_{BA} :

$$\begin{aligned} a_{AB}(\theta) &\equiv \frac{1}{L} \int A(\tilde{d}(\theta, \varphi, \theta', \varphi')) b_{AB}(\theta, \varphi, \theta', \varphi') dL(\theta', \varphi'), \\ a_{BA}(\theta) &\equiv \frac{1}{L} \int A(\tilde{d}(\theta, \varphi, \theta', \varphi')) b_{BA}(\theta, \varphi, \theta', \varphi') dL(\theta', \varphi') .. \end{aligned}$$

Because $a(\theta, \varphi)$ is independent of φ , its spherical harmonic coefficients a_l^m with nonzero m vanish. The definition of may be rewritten as

$$a_{AB}(\theta) = \rho_L 1_{C_A}(\theta, \varphi) \int_0^\pi \int_0^{2\pi} A(d(\theta, \varphi, \theta', \varphi')) 1_{C_B}(\theta', \varphi') \sin \theta' d\varphi' d\theta'.$$

The integral on the right-hand side depends only on θ , the dependence on φ is trivial.

To find its value, notice that it is equal to the spherical convolution $(A * 1_{C_B})(\theta, \varphi)$.

With the help of the formula (1.53), its spherical harmonic coefficients are simply

$$(A * 1_{C_B})_l^0 = \frac{\sqrt{4\pi}}{\sqrt{2l+1}} A_l^0 (1_{C_B})_l^0,$$

which means that, according to (1.50),

$$\int_0^\pi \int_0^{2\pi} A(d(\theta, \varphi, \theta', \varphi')) 1_{C_B}(\theta', \varphi') d\varphi' d\theta' = \sum_{l=0}^{\infty} \frac{\sqrt{4\pi}}{\sqrt{2l+1}} A_l^0 (1_{C_B})_l^0 Y_l^0(\theta, \varphi).$$

As a result, the expression for $a_{AB}(\theta)$ becomes

$$a_{AB}(\theta) = \rho_L \sum_{l=0}^{\infty} \frac{\sqrt{4\pi}}{\sqrt{2l+1}} A_l^0 (1_{C_B})_l^0 Y_l^0(\theta, \varphi) 1_{C_A}(\theta, \varphi),$$

or equivalently,

$$a_{AB}(\theta) = \frac{L}{\sqrt{4\pi}} \sum_{l=0}^{\infty} \left(\sum_{l'=0}^{\infty} \frac{1}{\sqrt{2l'+1}} A_{l'}^0 (1_{C_B})_{l'}^0 [Y_{l'}^0 1_{C_A}]_l^0 \right) Y_l^0(\theta, \varphi).$$

Analogously,

$$a_{BA}(\theta) = \frac{L}{\sqrt{4\pi}} \sum_{l=0}^{\infty} \left(\sum_{l'=0}^{\infty} \frac{1}{\sqrt{2l'+1}} A_{l'}^0 (1_{C_A})_{l'}^0 [Y_{l'}^0 1_{C_B}]_l^0 \right) Y_l^0(\theta, \varphi).$$

Adding the last two equations and comparing the result to (1.50) yields the following expression for the spherical harmonic coefficients of $a(\theta, \varphi)$:

$$a_l^0 = \frac{L}{\sqrt{4\pi}} \sum_{l'=0}^{\infty} \frac{A_{l'}^0}{\sqrt{2l'+1}} \left((1_{C_B})_{l'}^0 [Y_{l'}^0 1_{C_A}]_l^0 + (1_{C_A})_{l'}^0 [Y_{l'}^0 1_{C_B}]_l^0 \right).$$

The values of $(1_{C_A})_{l'}^0$, $(1_{C_B})_{l'}^0$, $[Y_{l'}^0 1_{C_A}]_l^0$, and $[Y_{l'}^0 1_{C_B}]_l^0$ were computed in earlier parts of this appendix.

A.9.4 Spherical harmonic expansion of $\tilde{g}_c(x)$

This result can be immediately applied (in the case of border costs) to the function $\tilde{g}_c(x) \equiv \int \tilde{G}_c(x, x') dL(x')$ defined in (1.23) as $\tilde{g}_c(x) \equiv \int \tilde{G}_c(x, x') dL(x') :$

$$\tilde{g}_{c,l} \equiv (\tilde{g}_c)_l^0 = \frac{1}{\sqrt{4\pi}} \sum_{l'=0}^{\infty} \frac{G_{c,l'}}{\sqrt{2l'+1}} \left((1_{C_B})_{l'}^0 [Y_{l'}^0 1_{C_A}]_l^0 + (1_{C_A})_{l'}^0 [Y_{l'}^0 1_{C_B}]_l^0 \right). \quad (\text{A.19})$$

Of course, due to rotational symmetry, $(\tilde{g}_c)_l^m = 0$ for $m \neq 0$. Analogously to the case of the circle, $(\hat{g}_c)_l^m = (\tilde{g}_c)_l^m$ for any l and m .

A.10 Derivation of the expression for T_l for the sphere

The spherical harmonic coefficients T_l^m of

$$T(\theta) = \left(\frac{1}{1 + 4\alpha^2 R^2 \sin^2 \frac{\theta}{2}} \right)^\delta = \left(\frac{1}{1 + 2\alpha^2 R^2 - 2\alpha^2 R^2 \cos \theta} \right)^\delta$$

can be computed using (1.51). For nonzero m T_l^m vanishes because $T(\theta)$ is independent of φ . For zero m , write $T_l \equiv T_l^0$. In this case (1.51) and (1.52) give

$$T_l = \int_0^\pi \int_0^{2\pi} T(\theta) Y_l^{0*}(\theta, \varphi) d\varphi \sin \theta d\theta = \sqrt{\pi} \sqrt{2l+1} \int_0^\pi T(\theta) P_l(\cos \theta) \sin \theta d\theta.$$

Performing the substitution $t \equiv \cos \theta$ in the integral gives

$$\begin{aligned} T_l &= \sqrt{\pi} \sqrt{2l+1} \int_0^\pi \frac{P_l(\cos \theta) \sin \theta d\theta}{(1 + 2\alpha^2 R^2 - 2\alpha^2 R^2 \cos \theta)^\delta} \\ &= \sqrt{\pi} \sqrt{2l+1} \int_{-1}^1 \frac{P_l(t) dt}{(1 + 2\alpha^2 R^2 - 2\alpha^2 R^2 t)^\delta}. \end{aligned}$$

As in the case of the circle, define $Z \equiv 1/\sqrt{1 + 4\alpha^2 R^2}$, which implies $2\alpha^2 R^2 = \frac{1-Z^2}{2Z^2}$.

$$T_l = \sqrt{\pi} \sqrt{2l+1} \left(\frac{2Z^2}{1-Z^2} \right)^\delta \int_{-1}^1 \frac{P_l(t)}{\left(\frac{1+Z^2}{1-Z^2} - t \right)^\delta} dt.$$

The value of the integral can be found in Gradshteyn and Ryzhik (2007), where equation 7.228 on p. 791 states that

$$\frac{1}{2} \Gamma(1+\mu) \int_{-1}^1 P_l(x) (z-x)^{-\mu-1} dx = (z^2-1)^{-\frac{\mu}{2}} e^{-i\pi\mu} Q_l^\mu(z).$$

With the replacement $\{\mu, z, x\} \rightarrow \left\{\delta - 1, \frac{1+Z^2}{1-Z^2}, t\right\}$ (which also means $z^2 - 1 \rightarrow \frac{4Z^2}{(1-Z^2)^2}$), this is

$$\int_{-1}^1 \frac{P_l(t)}{\left(\frac{1+Z^2}{1-Z^2} - t\right)^\delta} dt = \frac{2}{\Gamma(\delta)} \left(\frac{1-Z^2}{2Z}\right)^{\delta-1} e^{-i\pi(\delta-1)} Q_l^{\delta-1} \left(\frac{1+Z^2}{1-Z^2}\right).$$

An alternative form of the right-hand side may be found using Gradshteyn and Ryzhik (2007), p. 959, eq. 8.703,

$$Q_\nu^\mu(z) = \frac{e^{\mu\pi i} \Gamma(\nu + \mu + 1) \Gamma\left(\frac{1}{2}\right)}{2^{\nu+1} \Gamma\left(\nu + \frac{3}{2}\right)} (z^2 - 1)^{\frac{\mu}{2}} z^{-\nu-\mu-1} F\left(\frac{\nu + \mu + 2}{2}, \frac{\nu + \mu + 1}{2}, \nu + \frac{3}{2}; \frac{1}{z^2}\right).$$

Replacing $\{\mu, \nu, z\} \rightarrow \left\{\delta - 1, l, \frac{1+Z^2}{1-Z^2}\right\}$ and noting that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$,

$$\begin{aligned} Q_l^{\delta-1} \left(\frac{1+Z^2}{1-Z^2}\right) &= \frac{e^{(\delta-1)\pi i} \sqrt{\pi} \Gamma(l + \delta) (2Z)^{\delta-1} (1-Z^2)^{l+1}}{2^{l+1} \Gamma\left(l + \frac{3}{2}\right) (1+Z^2)^{l+\delta}} \\ &\quad \times F\left(\frac{l + \delta + 1}{2}, \frac{l + \delta}{2}, l + \frac{3}{2}; \left(\frac{1-Z^2}{1+Z^2}\right)^2\right). \end{aligned}$$

As a result, the integral can be rewritten as

$$\begin{aligned} &\int_{-1}^1 \frac{P_l(t) dt}{\left(\frac{1+Z^2}{1-Z^2} - t\right)^\delta} \\ &= \frac{\sqrt{\pi} \Gamma(l + \delta)}{2^l \Gamma(\delta) \Gamma\left(l + \frac{3}{2}\right)} \left(\frac{1-Z^2}{1+Z^2}\right)^{l+\delta} F\left(\frac{l + \delta + 1}{2}, \frac{l + \delta}{2}, l + \frac{3}{2}; \left(\frac{1-Z^2}{1+Z^2}\right)^2\right), \end{aligned}$$

and spherical harmonic coefficient T_l becomes

$$T_l = \frac{\pi \sqrt{2l+1} \Gamma(l + \delta) (2Z^2)^\delta (1-Z^2)^l}{2^l \Gamma(\delta) \Gamma\left(l + \frac{3}{2}\right) (1+Z^2)^{l+\delta}} F\left(\frac{l + \delta + 1}{2}, \frac{l + \delta}{2}, l + \frac{3}{2}; \left(\frac{1-Z^2}{1+Z^2}\right)^2\right). \quad (\text{A.20})$$

The Gauss hypergeometric function on the right-hand side may be further manipulated using several other identities. Gradshteyn and Ryzhik (2007), p. 1009, equation 9.134(2), reads

$$F(2\alpha, 2\alpha + 1 - \gamma, \gamma; z) = (1 + z)^{-2\alpha} F\left(\alpha, \alpha + \frac{1}{2}, \gamma; \frac{4z}{(1 + z)^2}\right).$$

Replacement $\{\alpha, \gamma, z\} \rightarrow \left\{\frac{l+\delta}{2}, l + \frac{3}{2}, \left(\frac{1-Z}{1+Z}\right)^2\right\}$ (which implies also $1 + z \rightarrow 2\frac{1+Z^2}{(1+Z)^2}$ and $\frac{4z}{(1+z)^2} \rightarrow \left(\frac{1-Z^2}{1+Z^2}\right)^2$) leads to

$$\begin{aligned} F\left(l + \delta, \delta - \frac{1}{2}, l + \frac{3}{2}; \left(\frac{1 - Z}{1 + Z}\right)^2\right) &= \frac{(1 + Z)^{2l+2\delta}}{2^{l+\delta} (1 + Z^2)^{l+\delta}} \\ &\times F\left(\frac{l + \delta}{2}, \frac{l + \delta + 1}{2}, l + \frac{3}{2}; \left(\frac{1 - Z^2}{1 + Z^2}\right)^2\right). \end{aligned} \quad (\text{A.21})$$

Equation 9.131(1) on p. 1008 of Gradshteyn and Ryzhik (2007) states that

$$F(\alpha, \beta, \gamma; z) = (1 - z)^{-\beta} F\left(\beta, \gamma - \alpha, \gamma; \frac{z}{z - 1}\right).$$

Replacing $\{\alpha, \beta, \gamma, z\} \rightarrow \left\{l + \delta, \delta - \frac{1}{2}, l + \frac{3}{2}, \left(\frac{1-Z}{1+Z}\right)^2\right\}$ (and consequently $1 - z \rightarrow \frac{4Z}{(1+Z)^2}$ and $\frac{z}{z-1} \rightarrow -\frac{(1-Z)^2}{4Z}$) gives

$$F\left(l + \delta, \delta - \frac{1}{2}, l + \frac{3}{2}; \left(\frac{1 - Z}{1 + Z}\right)^2\right) = \frac{(4Z)^{\frac{1}{2}-\delta}}{(1 + Z)^{1-2\delta}} F\left(\delta - \frac{1}{2}, \frac{3}{2} - \delta, l + \frac{3}{2}; -\frac{(1 - Z)^2}{4Z}\right). \quad (\text{A.22})$$

Equation 8.702 on p. 959 of Gradshteyn and Ryzhik (2007) reads

$$P_v^\mu(z) = \frac{1}{\Gamma(1 - \mu)} \left(\frac{z + 1}{z - 1}\right)^{\frac{\mu}{2}} F\left(-\nu, \nu + 1, 1 - \mu; \frac{1 - z}{2}\right).$$

Replacement $\{\mu, \nu, z\} \rightarrow \left\{-l - \frac{1}{2}, \delta - \frac{3}{2}, \frac{1+Z^2}{2Z}\right\}$ (and $\frac{z+1}{z-1} \rightarrow \left(\frac{1+Z}{1-Z}\right)^2$, $\frac{1-z}{2} \rightarrow -\frac{(1-Z)^2}{4Z}$) leads to

$$P_{\delta-\frac{3}{2}}^{-l-\frac{1}{2}}\left(\frac{1+Z^2}{2Z}\right) = \frac{1}{\Gamma(l+\frac{3}{2})} \left(\frac{1-Z}{1+Z}\right)^{l+\frac{1}{2}} F\left(\frac{3}{2}-\delta, \delta-\frac{1}{2}, l+\frac{3}{2}; -\frac{(1-Z)^2}{4Z}\right). \quad (\text{A.23})$$

The definition of the Gauss hypergeometric function (e.g. in Section 9.1 of Gradshteyn and Ryzhik (2007)) implies that the function is invariant under the exchange of its first two arguments. For this reason, (A.22) and (A.23) give

$$P_{\delta-\frac{3}{2}}^{-l-\frac{1}{2}}\left(\frac{1+Z^2}{2Z}\right) = \frac{(4Z)^{\delta-\frac{1}{2}}}{\Gamma(l+\frac{3}{2})} \frac{(1-Z)^{l+\frac{1}{2}}}{(1+Z)^{l+2\delta-\frac{1}{2}}} F\left(l+\delta, \delta-\frac{1}{2}, l+\frac{3}{2}; \left(\frac{1-Z}{1+Z}\right)^2\right).$$

This can be combined with (A.21) to give

$$P_{\delta-\frac{3}{2}}^{-l-\frac{1}{2}}\left(\frac{1+Z^2}{2Z}\right) = \frac{(4Z)^{\delta-\frac{1}{2}}}{\Gamma(l+\frac{3}{2})} \frac{(1-Z^2)^{l+\frac{1}{2}}}{2^{l+\delta}(1+Z^2)^{l+\delta}} F\left(\frac{l+\delta}{2}, \frac{l+\delta+1}{2}, l+\frac{3}{2}; \left(\frac{1-Z^2}{1+Z^2}\right)^2\right).$$

Recalling that F is symmetric in its first two arguments and substituting the last equation into (A.20) leads to the final result

$$T_l = 2\pi\sqrt{2l+1}(\delta)_l \frac{Z^{\frac{1}{2}-\delta}}{\sqrt{1-Z^2}} P_{\delta-\frac{3}{2}}^{-l-\frac{1}{2}}\left(\frac{1+Z^2}{2Z}\right). \quad (\text{A.24})$$

Here the Pochhammer symbol $(\delta)_l$ is defined as $\Gamma(l+\delta)/\Gamma(\delta) = \delta(\delta+1)\dots(\delta+l-1)$.

A.11 Relation to fields in anti de Sitter space

The parameter threshold discussed here has a counterpart in physics, namely the Breitenlohner and Freedman (1982a,b) bound that applies to fields in anti de Sitter space. The variables of the economic models with asymptotically power-law trade

costs share one important property with fields in anti de Sitter space, namely the behavior of their propagators at long distances. The relevant comparison here is between a d_s -dimensional economic model and fields in a $(d_s + 1)$ -dimensional anti de Sitter space, which has a d_s -dimensional boundary where exogenous changes can be introduced.

Scalar fields in $(d_s + 1)$ -dimensional anti de Sitter space have propagators that at large distances d scale like $d^{-2\Delta}$ for a definite parameter Δ , which depends on their mass. The minimum mass-squared that the stability of the system allows is given by the Breitenlohner-Freedman value of $-d_s^2 / (4R_{AdS}^2)$, where R_{AdS} is the curvature radius of the anti de Sitter space; see eq. (2.42) of Aharony, Gubser, Maldacena, Ooguri, and Oz (2000). Due to eq. (3.14) of Aharony, Gubser, Maldacena, Ooguri, and Oz (2000), this corresponds to $\Delta = d_s/2$.

In the economics situation of Section 1.4, the consumption part of the GDP propagator behaves at long distances like $d^{-2\delta}$, which means that δ can be thought of as the economics counterpart of Δ . Via this identification the physics relation $\Delta = d_s/2$ translates to the economics relation $\delta = d_s/2$, which is precisely the threshold where the qualitative behavior of the trade model changes.

Note that the explicit form of the propagator (3.42) of Aharony, Gubser, Maldacena, Ooguri, and Oz (2000) is the same as the consumption part (1.12) of the GDP propagator when the trade costs are $\tilde{\tau} = \left(1 + (\alpha d)^2\right)^{\rho/2}$. The same propagator (3.42) may be interpreted also from the point of view of the global anti de Sitter space, instead of the Poincaré coordinate patch perspective. When translated to the corresponding global anti de Sitter coordinates, the propagator acquires the same functional form as the consumption part (1.12) of the GDP propagator when the trade costs are $\tilde{\tau}(d) = \left(1 + 4\alpha^2 R^2 \sin^2(d/(2R))\right)^{\rho/2}$.

A.12 The impact of border costs in the large-space limit on the circle

This appendix derives the limit behavior (1.47) and (1.48) of the impact of border costs in the case of trade costs given by $\tilde{\tau}(d) = \left(1 + 4\alpha^2 R^2 \sin^2 \frac{d}{2R}\right)^{\frac{\ell}{2}}$. The corresponding expression for $G_{c,n}$ is given in (1.44). $G_{c,n}$ is independent of the sign of n , and is a decreasing function of $|n|$. The quantity Z , used frequently in this appendix, is defined as $Z \equiv 1/\sqrt{1 + 4\alpha^2 R^2}$.

A.12.1 Nonnegativity of $(-1)^{\frac{n}{2}} \tilde{g}_{c,n}$

Let us show that $(-1)^{\frac{n}{2}} \tilde{g}_{c,n} \geq 0$. The Fourier coefficients $\tilde{g}_{c,n}$ are given by (1.45). All $\tilde{g}_{c,n}$ with odd n vanish. Also, $\tilde{g}_{c,0} \geq 0$ because $\tilde{g}_c(\theta)$ is nonnegative for any θ . This means that it is sufficient to show that

$$\sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2} \quad (\text{A.25})$$

is nonpositive for even nonzero n . For this purpose, we can use the identity

$$\sum_{m=0}^{\infty} \frac{1}{(2m+1)^2 - n^2} = \frac{1}{2n} \sum_{m=0}^{\infty} \left(\frac{1}{2m+1-n} - \frac{1}{2m+1+n} \right) = \frac{1}{2n} \sum_{m=-\frac{n}{2}}^{\frac{n}{2}-1} \frac{1}{2m+1} = 0$$

to rewrite the term of interest (A.25) as

$$\sum_{m=0}^{\infty} \frac{LG_{c,2m+1} - LG_{c,n}}{(2m+1)^2 - n^2}.$$

If $2m+1 < n$, both the numerator and the denominator are negative. If $2m+1 > n$, they are both positive. This means that all contributions to the infinite sum are positive. We conclude that $(-1)^{\frac{n}{2}} \tilde{g}_{c,n} \geq 0$.

A.12.2 A lower bound on $-y_1^{(P)}\left(\frac{\pi}{2}\right)$

Plugging the expressions (1.41) for $y_{1,n}^{(P)}$ into the general formula for Fourier series expansion (1.37) gives

$$-\frac{y_1^{(P)}\left(\frac{\pi}{2}\right)}{y_0^{(P)}} = -\frac{1}{y_0^{(P)}} \sum_{n=-\infty}^{\infty} y_{1,n}^{(P)} e^{in\frac{\pi}{2}} = \frac{\sigma}{\sigma-1} \tilde{g}_{c,0} + \frac{2\sigma-1}{\sigma-1} \sum_{n \text{ even nonzero}} \frac{(-1)^{\frac{n}{2}} \tilde{g}_{c,n}}{1 + (\sigma-1) LG_{c,n}}. \quad (\text{A.26})$$

We know from Subsection 1.5.4 that $LG_{c,0} = 1/\sigma$ and from Subsection A.12.1 that $(-1)^{\frac{n}{2}} \tilde{g}_{c,n} \geq 0$. Equation (A.26) combined with $0 \leq LG_{c,n} \leq 1/\sigma$ and $(-1)^{\frac{n}{2}} \tilde{g}_{c,n} \geq 0$ implies

$$-\frac{y_1^{(P)}\left(\frac{\pi}{2}\right)}{y_0^{(P)}} \geq \frac{\sigma}{\sigma-1} \sum_{n \text{ even}} (-1)^{\frac{n}{2}} \tilde{g}_{c,n}.$$

Taking into account the symmetry between the countries, the fact that $\int G_c(\theta) dL(\theta) = LG_{c,0} = 1/\sigma$, and the definition (1.24) of \tilde{g}_c , we see that

$$\sum_{n \text{ even}} (-1)^{\frac{n}{2}} \tilde{g}_{c,n} = \frac{1}{2\sigma}. \quad (\text{A.27})$$

This leads to the following lower bound on $-y_1^{(P)}\left(\frac{\pi}{2}\right)$:

$$-\frac{y_1^{(P)}\left(\frac{\pi}{2}\right)}{y_0^{(P)}} \geq \frac{1}{2} \frac{1}{\sigma-1}. \quad (\text{A.28})$$

A.12.3 An upper bound on $-y_1^{(P)}\left(\frac{\pi}{2}\right)$

Equation (A.26) may be used to derive also an upper bound on $-y_1^{(P)}\left(\frac{\pi}{2}\right)$. Using $LG_{c,n} \geq 0$ and $(-1)^{\frac{n}{2}} \tilde{g}_{c,n} \geq 0$ leads to

$$-\frac{y_1^{(P)}\left(\frac{\pi}{2}\right)}{y_0^{(P)}} \leq -\tilde{g}_{c,0} + \frac{2\sigma-1}{\sigma-1} \sum_{n \text{ even}} (-1)^{\frac{n}{2}} \tilde{g}_{c,n}.$$

Omitting the first term on the right-hand side and simplifying the second term using (A.27) gives

$$-\frac{y_1^{(P)}\left(\frac{\pi}{2}\right)}{y_0^{(P)}} \leq \frac{2\sigma-1}{2\sigma} \frac{1}{\sigma-1}. \quad (\text{A.29})$$

A.12.4 A lower bound on $-\lim_{R \rightarrow \infty} y_1^{(P)}(0)$ for $\delta < \frac{1}{2}$

This subsection contains the derivation of a lower bound on $-\lim_{R \rightarrow \infty} y_1^{(P)}(0)$. To simplify notation, the limit symbol $\lim_{R \rightarrow \infty}$ will be omitted, but its presence is implicitly understood.

The asymptotic form of $G_{c,n}$ for $\delta < \frac{1}{2}$

For arbitrary R , the expression (1.44) for $G_{c,n}$ is

$$G_{c,n} = \frac{1}{\sigma L} \frac{(-1)^n}{(1-\delta)_n} \frac{P_{\delta-1}^n\left(\frac{1+Z^2}{2Z}\right)}{P_{\delta-1}\left(\frac{1+Z^2}{2Z}\right)}.$$

In the large R limit the expression simplifies to

$$G_{c,n} = \frac{1}{\sigma L} \frac{(\delta)_n}{(1-\delta)_n}. \quad (\text{A.30})$$

This asymptotic form can be verified using the definition of the Pochhammer symbol

$$(\delta)_n \equiv \frac{\Gamma(\delta+n)}{\Gamma(\delta)} = (-1)^n \frac{\Gamma(1-\delta)}{\Gamma(1-\delta-n)},$$

and the equation 8.766(1) on p. 971 of Gradshteyn and Ryzhik (2007), which states that for $|z| \gg 1$,

$$P_\nu^\mu(z) = \left\{ \frac{2^\nu \Gamma(\nu + \frac{1}{2})}{\sqrt{\pi} \Gamma(\nu - \mu + 1)} z^\nu + \frac{\Gamma(-\nu - \frac{1}{2})}{2^{\nu+1} \sqrt{\pi} \Gamma(-\nu - \mu)} z^{-\nu-1} \right\} \left(1 + O\left(\frac{1}{z^2}\right) \right).$$

The Fourier series expansion of $y_1^{(P)}(0)$

Using the expressions (1.46) for $y_{1,n}^{(P)}$ in the general formula for Fourier series expansion (1.38) gives

$$\begin{aligned} -\frac{y_1^{(P)}(0)}{y_0^{(P)}} &= \frac{1}{2} \frac{1}{\sigma - 1} - \frac{4}{\pi^2} \frac{\sigma}{\sigma - 1} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2} \\ &\quad - \frac{4}{\pi^2} \frac{2\sigma - 1}{\sigma - 1} \sum_{n \text{ even nonzero}} \frac{(-1)^{\frac{n}{2}}}{1 + (\sigma - 1) LG_{c,n}} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2}. \end{aligned}$$

This relation may be rewritten as

$$\begin{aligned} -\frac{y_1^{(P)}(0)}{y_0^{(P)}} &= -\left(\frac{1}{2\sigma} - \frac{4}{\pi^2} \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2} LG_{c,2m+1} \right) \\ &\quad + \frac{2\sigma - 1}{\sigma - 1} \left(\frac{1}{2\sigma} - \frac{4}{\pi^2} \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2} LG_{c,2m+1} \right) \\ &\quad - \frac{4}{\pi^2} \frac{2\sigma - 1}{\sigma - 1} \sum_{n \text{ even nonzero}} (-1)^{\frac{n}{2}} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2} \\ &\quad + \frac{4}{\pi^2} \frac{2\sigma - 1}{\sigma - 1} \sum_{n \text{ even nonzero}} \frac{(-1)^{\frac{n}{2}} (\sigma - 1) LG_{c,n}}{1 + (\sigma - 1) LG_{c,n}} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2}. \end{aligned}$$

The first term on the right-hand side is just $-\tilde{g}_{c,0}$ due to (1.45). The terms on the second and third line add up to $\frac{2\sigma-1}{\sigma-1} \tilde{g}_c(0)$, as implied by the Fourier series expansion

formula (1.37) with the Fourier coefficients (1.45):

$$\tilde{g}_c(0) = \frac{1}{2\sigma} - \frac{4}{\pi^2} \sum_{n \text{ even}} (-1)^{\frac{n}{2}} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2}. \quad (\text{A.31})$$

The Fourier series expansion of $y_1^{(P)}(0)$ can now be written as

$$\begin{aligned} -\frac{y_1^{(P)}(0)}{y_0^{(P)}} &= -\tilde{g}_{c,0} + \frac{2\sigma-1}{\sigma-1} \tilde{g}_c(0) \\ &+ \frac{4}{\pi^2} \frac{2\sigma-1}{\sigma-1} \sum_{\substack{n \text{ even} \\ \text{nonzero}}} \frac{(-1)^{\frac{n}{2}} (\sigma-1) LG_{c,n}}{1 + (\sigma-1) LG_{c,n}} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2} \end{aligned} \quad (\text{A.32})$$

The rest of this subsection analyzes the properties of the terms on the right-hand side in the large-space limit. The analysis then leads to an asymptotic lower bound on $-y_1^{(P)}(0)$.

Evaluating $\lim_{R \rightarrow \infty} \tilde{g}_c(0)$

The expression (A.31) for $\tilde{g}_c(0)$ with the asymptotic form (A.30) of $G_{c,n}$ is

$$\sigma \tilde{g}_c(0) = \frac{1}{2} - \frac{4}{\pi^2} \sum_{n \text{ even}} (-1)^{\frac{n}{2}} \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2 - n^2} \frac{(\delta)_{2m+1}}{(1-\delta)_{2m+1}}.$$

The identity

$$\begin{aligned} &\sum_{n \text{ even}} \frac{4(-1)^{\frac{n}{2}}}{\pi^2} \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2 - n^2} \frac{(\delta)_{2m+1}}{(1-\delta)_{2m+1}} \\ &= \frac{1}{2} - 2 \frac{\Gamma(1-\delta)}{\sqrt{\pi} \Gamma(\frac{1}{2}-\delta)} F\left(\frac{1}{2}, 1-\delta; \frac{3}{2}; -1\right) \end{aligned}$$

then leads to the following compact result for $\tilde{g}_c(0)$:

$$\sigma \tilde{g}_c(0) = 2 \frac{\Gamma(1-\delta)}{\sqrt{\pi} \Gamma(\frac{1}{2}-\delta)} F\left(\frac{1}{2}, 1-\delta; \frac{3}{2}; -1\right). \quad (\text{A.34})$$

An alternative way of evaluating of $\lim_{R \rightarrow \infty} \tilde{g}_c(0)$

An alternative way of deriving (A.34) is to work directly with the definition (1.24) of $\tilde{g}_c(\theta)$, which implies

$$\tilde{g}_c(0) = 2\rho_L \int_{\frac{\pi}{2}}^{\pi} G_c(\theta) d\theta.$$

Substituting the expression (1.42) for $G_c(\theta)$, we get

$$\tilde{g}_c(0) = \frac{1}{\sigma} \frac{\int_{\frac{\pi}{2}}^{\pi} T(\theta) d\theta}{\int_0^{\pi} T(\theta) d\theta}.$$

For the functional form (1.43) of $T(\theta)$ this is

$$\tilde{g}_c(0) = \frac{1}{\sigma} \frac{\int_{\frac{\pi}{2}}^{\pi} (Z^2 \cos^2 \frac{\theta}{2} + \sin^2 \frac{\theta}{2})^{-\delta} d\theta}{\int_0^{\pi} (Z^2 \cos^2 \frac{\theta}{2} + \sin^2 \frac{\theta}{2})^{-\delta} d\theta}.$$

The large-space limit $R \rightarrow \infty$ corresponds to $Z \rightarrow 0_+$, and in this limit

$$\tilde{g}_c(0) = \frac{1}{\sigma} \frac{\int_{\frac{\pi}{2}}^{\pi} \sin^{-2\delta} \frac{\theta}{2} d\theta}{\int_0^{\pi} \sin^{-2\delta} \frac{\theta}{2} d\theta} = \frac{1}{\sigma} \frac{\int_{\frac{\pi}{2}}^{\pi} (1 - \cos \theta)^{-\delta} d\theta}{\int_0^{\pi} (1 - \cos \theta)^{-\delta} d\theta}.$$

To find an explicit expression for the integrals, we can use the substitution $t \equiv$

$$\frac{1+\cos \theta}{2}, d\theta = -\frac{1}{2} t^{-\frac{1}{2}} (1-t)^{-\frac{1}{2}} dt,$$

$$\tilde{g}_c(0) = \frac{1}{\sigma} \frac{\int_0^{\frac{1}{2}} (1-t)^{-\delta-\frac{1}{2}} t^{-\frac{1}{2}} dt}{\int_0^1 (1-t)^{-\delta-\frac{1}{2}} t^{-\frac{1}{2}} dt} = \frac{1}{\sigma} \frac{B_{\frac{1}{2}}(\frac{1}{2}, \frac{1}{2}-\delta)}{B_1(\frac{1}{2}, \frac{1}{2}-\delta)}. \quad (\text{A.35})$$

The second equality⁵ here follows from the definition of the incomplete beta function,⁶

$$B_x(p, q) = \int_0^x t^{p-1} (1-t)^{q-1} dt.$$

This special function should not be confused with the primary impact function (1.19). The result (A.35) for $\tilde{g}_c(0)$ matches the expression (A.34) derived by summing up the infinite series.

Evaluating $\lim_{R \rightarrow \infty} \tilde{g}_{c,0}$

Now let us look at $\tilde{g}_{c,0}$. The expression for $\tilde{g}_{c,0}$ in (1.45) with the asymptotic form (A.30) of $G_{c,n}$ becomes

$$\sigma \tilde{g}_{c,0} = \frac{1}{2} - \frac{4}{\pi^2} \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2} \frac{(\delta)_{2m+1}}{(1-\delta)_{2m+1}}.$$

The sum can be expressed in terms of the generalized hypergeometric function ${}_5F_4$,

$$\sigma \tilde{g}_{c,0} = \frac{1}{2} + \frac{4}{\pi^2} \frac{\delta}{1-\delta} {}_5F_4 \left(\frac{1}{2}, \frac{1}{2}, 1, \frac{1}{2} + \frac{\delta}{2}, 1 + \frac{\delta}{2}; \frac{3}{2}, \frac{3}{2}, 1 - \frac{\delta}{2}, \frac{3}{2} - \frac{\delta}{2}; 1 \right).$$

Positivity of the last term

Consider now the last term in (A.32):

$$\frac{4}{\pi^2} \frac{2\sigma-1}{\sigma-1} \sum_{n \text{ even nonzero}} \frac{(-1)^{\frac{n}{2}} (\sigma-1) LG_{c,n}}{1 + (\sigma-1) LG_{c,n}} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2}.$$

⁵ Note that $B_1\left(\frac{1}{2}, \frac{1}{2} - \delta\right)$ may be written in terms of the (complete) beta function as $B\left(\frac{1}{2}, \frac{1}{2} - \delta\right)$ or in terms of the gamma function as $\sqrt{\pi} \Gamma\left(\frac{1}{2} - \delta\right) / \Gamma(1 - \delta)$. $B_{\frac{1}{2}}\left(\frac{1}{2}, \frac{1}{2} - \delta\right)$ can be expressed in terms of the Gauss hypergeometric function as $\sqrt{2} F\left(\frac{1}{2}, \frac{1}{2} + \delta; \frac{3}{2}; \frac{1}{2}\right)$ or as $2F\left(\frac{1}{2}, 1 - \delta; \frac{3}{2}; -1\right)$.

⁶ See, for example, equation eq. 8.391 on p. 910 of Gradshteyn and Ryzhik (2007).

With $G_{c,n}$ given by (A.30), the inner sum can be evaluated explicitly. For n even and nonzero,

$$\begin{aligned}
& \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2 - n^2} \frac{(\delta)_{2m+1}}{(1-\delta)_{2m+1}} \\
&= -\frac{\delta}{2(1-\delta)} \frac{1}{n(n^2-1)} \\
& \quad \left((n-1) {}_4F_3 \left(1, \frac{1+n}{2}, \frac{1+\delta}{2}, \frac{1+\delta}{2}; \frac{3+n}{2}, 1-\frac{\delta}{2}, \frac{3}{2}-\frac{\delta}{2}; 1 \right) \right. \\
& \quad \left. + (n+1) {}_4F_3 \left(1, \frac{1-n}{2}, \frac{1+\delta}{2}, \frac{1+\delta}{2}; \frac{3-n}{2}, 1-\frac{\delta}{2}, \frac{3}{2}-\frac{\delta}{2}; 1 \right) \right).
\end{aligned}$$

Let us now restrict attention to $|n|$ even and nonzero. The expression $\sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2}$ is negative, and its absolute value is a decreasing function of $|n|$. Note also that $\frac{(\sigma-1)LG_{c,n}}{1+(\sigma-1)LG_{c,n}}$ is a positive decreasing function of $|n|$. The factor $(-1)^{\frac{n}{2}}$ is just an alternating sign, which is negative for the lowest terms, i.e. for $|n| = 2$. These facts imply that the last term in (A.32) is positive. Omitting this term in (A.32) leads to a lower bound on $-y_1^{(P)}(0)$.

$$-\frac{y_1^{(P)}(0)}{y_0^{(P)}} \geq -\tilde{g}_{c,0} + \frac{2\sigma-1}{\sigma-1} \tilde{g}_c(0).$$

Noting that $\frac{2\sigma-1}{\sigma-1} \geq 2$ and $\tilde{g}_c(0) \geq 0$, this implies the weaker bound

$$-\frac{y_1^{(P)}(0)}{y_0^{(P)}} \geq 2\tilde{g}_c(0) - \tilde{g}_{c,0}.$$

The resulting lower bound on $\lim_{R \rightarrow \infty} y_1^{(P)}(0)$

Substituting the expressions for $\tilde{g}_{c,0}$ and $\tilde{g}_c(0)$ into the last inequality, we get

$$-\frac{\sigma y_1^{(P)}(0)}{y_0^{(P)}} = \frac{4\Gamma(1-\delta)}{\sqrt{\pi}\Gamma(\frac{1}{2}-\delta)} F\left(\frac{1}{2}, 1-\delta; \frac{3}{2}; -1\right) - \frac{1}{2} - \frac{4}{\pi^2} \frac{\delta}{1-\delta} {}_5F_4\left(\frac{1}{2}, \frac{1}{2}, 1, \frac{1}{2} + \frac{\delta}{2}, 1 + \frac{\delta}{2}; \frac{3}{2}, \frac{3}{2}, 1 - \frac{\delta}{2}, \frac{3}{2} - \frac{\delta}{2}; 1\right).$$

The function on the right-hand side is concave, takes value of $\frac{1}{2}$ at $\delta = 0$, and vanishes at $\delta = \frac{1}{2}$. It is therefore never smaller than $\frac{1}{2} - \delta$. This leads to the result

$$-\frac{y_1^{(P)}(0)}{y_0^{(P)}} \geq \frac{\frac{1}{2} - \delta}{\sigma}. \quad (\text{A.36})$$

A.12.5 An alternative derivation of the bounds (A.28) and (A.29) in the large-space limit

for $\delta > \frac{1}{2}$

The large space limit of $G_{c,n}$ for $\delta > \frac{1}{2}$

Consider again the expression (1.44) for $G_{c,n}$:

$$G_{c,n} = \frac{1}{\sigma L} \frac{(-1)^n P_{\delta-1}^n\left(\frac{1+Z^2}{2Z}\right)}{(1-\delta)_n P_{\delta-1}\left(\frac{1+Z^2}{2Z}\right)},$$

with the goal of understanding the $R \rightarrow \infty$ (i.e. $Z \rightarrow 0_+$) limit when $\delta > \frac{1}{2}$. The definition⁷ of $P_\nu^\mu(z)$ is

$$P_\nu^\mu(z) = \left(\frac{z+1}{z-1}\right)^{\frac{\mu}{2}} {}_2\tilde{F}_1\left(-\nu, \nu+1; 1-\mu; \frac{1-z}{2}\right),$$

⁷ See <http://functions.wolfram.com/07.09.02.0001.01>

where ${}_2\tilde{F}_1$ is the regularized hypergeometric function. In order to apply this relation directly to the expression for $G_{c,n}$, we need to choose

$$z = \frac{1+Z^2}{2Z}, \quad \frac{1-z}{2} = -\frac{(1-Z)^2}{4Z}, \quad \frac{z+1}{z-1} = \left(\frac{1+Z}{1-Z}\right)^2.$$

The Fourier coefficients $G_{c,n}$ become

$$G_{c,n} = \frac{1}{\sigma L} \frac{(-1)^n}{(1-\delta)_n} \left(\frac{1+Z}{1-Z}\right)^n \frac{{}_2\tilde{F}_1\left(1-\delta, \delta; 1-n; -\frac{(1-Z)^2}{4Z}\right)}{{}_2\tilde{F}_1\left(1-\delta, \delta; 1; -\frac{(1-Z)^2}{4Z}\right)}.$$

Since $G_{c,n} = G_{c,-n}$, we know that the expression on the right-hand side does not depend on the sign of n . It is therefore sufficient to focus on non-positive n . In this case one can use the ordinary hypergeometric function ${}_2F_1$ instead of the regularized one. These functions are related⁸ by ${}_2\tilde{F}_1(a, b, c, x) \equiv {}_2F_1(a, b, c, x) / \Gamma(c)$, where $\Gamma(x)$ is the gamma function.

$$G_{c,n} = \frac{1}{\sigma L} \frac{(-1)^n}{(1-\delta)_n} \left(\frac{1+Z}{1-Z}\right)^n \frac{1}{\Gamma(1-n)} \frac{{}_2F_1\left(1-\delta, \delta; 1-n; -\frac{(1-Z)^2}{4Z}\right)}{{}_2F_1\left(1-\delta, \delta; 1; -\frac{(1-Z)^2}{4Z}\right)}.$$

For convenience, define the rescaled Fourier mode number⁹ ω as

$$\omega \equiv -Zn.$$

⁸ See <http://functions.wolfram.com/07.24.26.0003.01>

⁹ Not to be confused with the notation for different varieties of goods.

In terms of ω , the Fourier coefficients of G_c are

$$G_{c, \frac{\omega}{Z}} = \frac{1}{\sigma L} \left(\frac{1+Z}{1-Z} \right)^{-\frac{\omega}{Z}} \times \frac{(-1)^{\frac{\omega}{Z}}}{(1-\delta)_{-\frac{\omega}{Z}}} \frac{Z^{\delta-1}}{\Gamma(1+\frac{\omega}{Z})} \times \frac{{}_2F_1\left(1-\delta, \delta; 1+\frac{\omega}{Z}; -\frac{(1-Z)^2}{4Z}\right)}{Z^{\delta-1} {}_2F_1\left(1-\delta, \delta; 1; -\frac{(1-Z)^2}{4Z}\right)}. \quad (\text{A.37})$$

The right-hand side has been split into three factors. Let us look at the $Z \rightarrow 0_+$ limit (holding ω fixed) of each of them separately.

The first factor in (A.37) has a very simple limit. The definition of the exponential function implies

$$\lim_{Z \rightarrow 0} \frac{1}{\sigma L} \left(\frac{1+Z}{1-Z} \right)^{-\frac{\omega}{Z}} = \frac{1}{\sigma L} e^{-2\omega}.$$

To evaluate the limit of the second factor in (A.37), recall the definition of the Pochhammer symbol

$$(1-\delta)_{-\frac{\omega}{Z}} \equiv \frac{\Gamma(1-\delta-\frac{\omega}{Z})}{\Gamma(1-\delta)}.$$

With the help of the gamma function identity¹⁰ $\Gamma(1-x)\Gamma(x)\sin\pi x = \pi$, this is

$$(1-\delta)_{-\frac{\omega}{Z}} = \frac{1}{\Gamma(1-\delta)} \frac{1}{\Gamma(\delta+\frac{\omega}{Z})} \frac{\pi}{\sin\pi(\delta+\frac{\omega}{Z})} = (-1)^{-\frac{\omega}{Z}} \frac{1}{\Gamma(1-\delta)} \frac{1}{\Gamma(\delta+\frac{\omega}{Z})} \frac{\pi}{\sin\pi\delta},$$

where the second equality follows from the periodicity properties of the sine function with $\frac{\omega}{Z} \in \mathbb{Z}$. The desired limit is

$$\lim_{Z \rightarrow 0_+} \frac{(-1)^{-\frac{\omega}{Z}}}{(1-\delta)_{-\frac{\omega}{Z}}} \frac{Z^{\delta-1}}{\Gamma(1+\frac{\omega}{Z})} = \Gamma(1-\delta) \frac{\sin\pi\delta}{\pi} \lim_{Z \rightarrow 0_+} \frac{Z^{\delta-1} \Gamma(\delta+\frac{\omega}{Z})}{\Gamma(1+\frac{\omega}{Z})} = \Gamma(1-\delta) \frac{\sin\pi\delta}{\pi} \omega^{\delta-1}.$$

The second equality may be verified using Stirling's formula.

To find the limit of the third factor in (A.37), we need two identities. The first

¹⁰ See, for example, equation 8.334(3) on p. 896 of Gradshteyn and Ryzhik (2007).

one¹¹ reads

$$\lim_{c \rightarrow \infty} {}_2F_1 \left(a, a - b + 1; c; 1 - \frac{c}{z} \right) = z^a U(a, b, z).$$

The correct version of the second one¹² is

$$U(a, 2a, z) = \frac{1}{\sqrt{\pi}} z^{\frac{1}{2}-a} e^{\frac{z}{2}} K_{a-\frac{1}{2}} \left(\frac{z}{2} \right).$$

The function U is the confluent hypergeometric function of the second kind, and K is the modified Bessel function of the second kind. Combining the two identities, and using $K_\nu(x) = K_{-\nu}(x)$ yields

$$\lim_{c \rightarrow \infty} {}_2F_1 \left(a, 1 - a; c; 1 - \frac{c}{z} \right) = \frac{1}{\sqrt{\pi}} z^{\frac{1}{2}} e^{\frac{z}{2}} K_{\frac{1}{2}-a} \left(\frac{z}{2} \right).$$

The choice

$$a = 1 - \delta, \quad c = 1 + \frac{\omega}{Z}, \quad z = 4\omega$$

in this formula leads to

$$\begin{aligned} \lim_{Z \rightarrow 0_+} {}_2F_1 \left(1 - \delta, \delta; 1 + \frac{\omega}{Z}; -\frac{(1 - Z)^2}{4Z} \right) &= \lim_{Z \rightarrow 0_+} {}_2F_1 \left(1 - \delta, \delta; 1 + \frac{\omega}{Z}; \frac{4\omega - 1}{4\omega} - \frac{1}{4Z} \right) \\ &= \frac{2}{\sqrt{\pi}} e^{2\omega} \omega^{\frac{1}{2}} K_{\delta-\frac{1}{2}}(2\omega). \end{aligned}$$

For the denominator of the third factor in (A.37), asymptotic properties of hypergeo-

¹¹ Available at <http://functions.wolfram.com/07.33.09.0001.01>

¹² Available at <http://functions.wolfram.com/07.33.03.0007.01>

Note that the graphical version of the formula is wrong on the website (as of April 2012), but its Mathematica version is correct.

metric functions imply

$$\lim_{Z \rightarrow 0_+} Z^{\delta-1} {}_2F_1 \left(1 - \delta, \delta; 1; -\frac{(1-Z)^2}{4Z} \right) = 4^{1-\delta} \frac{\Gamma(2\delta-1)}{\Gamma^2(\delta)}.$$

We have

$$\lim_{Z \rightarrow 0_+} \frac{{}_2F_1 \left(1 - \delta, \delta; 1 + \frac{\omega}{Z}; -\frac{(1-Z)^2}{4Z} \right)}{Z^{\delta-1} {}_2F_1 \left(1 - \delta, \delta; 1; -\frac{(1-Z)^2}{4Z} \right)} = \frac{\frac{2}{\sqrt{\pi}} e^{2\omega} \omega^{\frac{1}{2}} K_{\delta-\frac{1}{2}}(2\omega)}{4^{1-\delta} \frac{\Gamma(2\delta-1)}{\Gamma^2(\delta)}}$$

Applying these three results to (A.37) gives

$$G_{c, \frac{\omega}{Z}} = \frac{1}{\sigma L} \frac{2^{2\delta-1} \Gamma(\delta)}{\sqrt{\pi} \Gamma(2\delta-1)} \Gamma(1-\delta) \Gamma(\delta) \frac{\sin \pi \delta}{\pi} \omega^{\delta-\frac{1}{2}} K_{\delta-\frac{1}{2}}(2\omega).$$

Using the gamma function identities¹³

$$\Gamma(1-\delta) \Gamma(\delta) \frac{\sin \pi \delta}{\pi} = 1$$

and

$$\frac{2^{2\delta-1} \Gamma(\delta)}{\sqrt{\pi} \Gamma(2\delta-1)} = \frac{2}{\Gamma(\delta - \frac{1}{2})},$$

the Fourier coefficients simplify to

$$G_{c, \frac{\omega}{Z}} = \frac{1}{\sigma L} \frac{2}{\Gamma(\delta - \frac{1}{2})} \omega^{\delta-\frac{1}{2}} K_{\delta-\frac{1}{2}}(2\omega).$$

We conclude that for arbitrary ω , in the large-space limit the Fourier coefficients become

$$G_{c, \frac{\omega}{Z}} = \frac{1}{\sigma L} \frac{2}{\Gamma(\delta - \frac{1}{2})} |\omega|^{\delta-\frac{1}{2}} K_{\delta-\frac{1}{2}}(2|\omega|). \quad (\text{A.38})$$

Notice that the properties of the modified Bessel function of the second kind

¹³ See, for example, equations 8.334(3) and 8.335(1) on p. 896 of Gradshteyn and Ryzhik (2007).

imply that the expression on the right-hand side approaches $1/(\sigma L)$ when $\omega \rightarrow 0$, as expected from the general relation $\sigma LG_{c,0} = 1$.

The limit of $y^{(P)}(\frac{\pi}{2})$ for $\delta > \frac{1}{2}$

The general formula for Fourier series expansion (1.38) together with the expressions (1.46) for $y_{1,n}^{(P)}$ implies

$$\begin{aligned} -(\sigma - 1) \frac{y_1^{(P)}(\frac{\pi}{2})}{y_0^{(P)}} &= \frac{1}{2} - \frac{4}{\pi^2} \sum_{m=0}^{\infty} \frac{\sigma LG_{c,2m+1}}{(2m+1)^2} \\ &\quad - \sum_{n \text{ even nonzero}} \frac{2\sigma - 1}{1 + (\sigma - 1) LG_{c,n}} \frac{4}{\pi^2} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2}. \end{aligned}$$

The $R \rightarrow \infty$ limit, or equivalently the $Z \rightarrow 0_+$ limit, of the first line on the right-hand side vanishes.¹⁴ We have

$$-(\sigma - 1) \lim_{R \rightarrow \infty} \frac{y_1^{(P)}(\frac{\pi}{2})}{y_0^{(P)}} = - \lim_{R \rightarrow \infty} \sum_{n \text{ even nonzero}} \frac{2\sigma - 1}{1 + (\sigma - 1) LG_{c,n}} \frac{4}{\pi^2} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2}.$$

In the $Z \rightarrow 0_+$ limit the sums can be faithfully approximated by integrals. Symbolically,

$$\begin{aligned} \sum_{m=0}^{\infty} f(2m+1) &\rightarrow \frac{1}{4Z} \int f(\omega') d\omega', \text{ for an even function } f, \\ \sum_{n \text{ even nonzero}} &\rightarrow \frac{1}{2Z} \int d\omega. \end{aligned}$$

¹⁴ As $Z \rightarrow 0_+$, the coefficients $G_{c,2m+1}$ approach $1/(\sigma L)$, and the convergence is uniform in the appropriate sense. Also, $\sum_{m=0}^{\infty} (2m+1)^{-2} = \pi^2/8$.

More precisely, the integral over ω' should be taken in the sense of the Cauchy principal value (denoted p.v.). This gives

$$-(\sigma - 1) \lim_{R \rightarrow \infty} \frac{y_1^{(P)}\left(\frac{\pi}{2}\right)}{y_0^{(P)}} = -\frac{1}{2\pi^2} \lim_{R \rightarrow \infty} \int \frac{2\sigma - 1}{1 + (\sigma - 1) LG_{c, \frac{\omega}{Z}}} \text{p.v.} \int \frac{LG_{c, \frac{\omega'}{Z}}}{\omega'^2 - \omega^2} d\omega' d\omega.$$

Using the algebraic relation

$$\frac{1}{\omega'^2 - \omega^2} = -\frac{1}{2\omega} \frac{1}{\omega - \omega'} - \frac{1}{2\omega} \frac{1}{\omega + \omega'}$$

and the explicit expression (A.38) for $G_{c, \frac{\omega'}{Z}}$, we obtain

$$\begin{aligned} &-(\sigma - 1) \lim_{R \rightarrow \infty} \frac{y_1^{(P)}\left(\frac{\pi}{2}\right)}{y_0^{(P)}} = \frac{1}{\sigma L} \frac{1}{\pi^2 \Gamma\left(\delta - \frac{1}{2}\right)} \\ &\times \int \frac{2\sigma - 1}{1 + \frac{\sigma-1}{\sigma} \frac{2}{\Gamma\left(\delta - \frac{1}{2}\right)}} \frac{1}{\omega} \text{p.v.} \int \frac{|\omega'|^{\delta - \frac{1}{2}} K_{\delta - \frac{1}{2}}(2|\omega'|)}{\omega - \omega'} d\omega' d\omega. \end{aligned}$$

The integral over ω' can be evaluated explicitly,

$$\begin{aligned} &\text{p.v.} \int \frac{1}{\omega - \omega'} |\omega'|^{\delta - \frac{1}{2}} K_{\delta - \frac{1}{2}}(2|\omega'|) d\omega' \\ &= \frac{1}{4\pi} \frac{\omega}{|\omega|} |\omega|^{\delta - \frac{1}{2}} G_{4,2}^{4,2} \left(\omega^2 \left| \begin{array}{c} \frac{1}{4}(1 - 2\delta), \frac{1}{4}(3 - 2\delta) \\ \frac{1}{4}(1 - 2\delta), \frac{1}{4}(1 - 2\delta), \frac{1}{4}(3 - 2\delta), -\frac{1}{4}(1 - 2\delta) \end{array} \right. \right) \\ &+ \frac{\pi\omega|\omega|^{\delta - \frac{3}{2}}}{2} G_{6,4}^{4,2} \left(\omega^2 \left| \begin{array}{c} \frac{1}{4}(1 - 2\delta), \frac{1}{4}(3 - 2\delta), \frac{1}{2}(1 - \delta), -\frac{1}{2}\delta \\ \frac{1}{4}(1 - 2\delta), \frac{1}{4}(1 - 2\delta), \frac{1}{4}(3 - 2\delta), -\frac{1}{4}(1 - 2\delta), \frac{1}{2}(1 - \delta), -\frac{1}{2}\delta \end{array} \right. \right), \end{aligned}$$

where $G_{p,q}^{m,n}$ is the Meijer G-function. The outer integral over ω most likely does not lead to a closed form expression, but it can easily be evaluated numerically.

Since $\frac{1}{\omega} \text{p.v.} \int \frac{|\omega'|^{\delta - \frac{1}{2}} K_{\delta - \frac{1}{2}}(2|\omega'|)}{\omega - \omega'} d\omega'$ is positive, and $\frac{2}{\Gamma\left(\delta - \frac{1}{2}\right)} |\omega|^{\delta - \frac{1}{2}} K_{\delta - \frac{1}{2}}(2|\omega|) \in (0, 1)$,

one can immediately write the following bound on $y_1^{(P)}\left(\frac{\pi}{2}\right)$ in the large-space limit.

$$\begin{aligned} & \frac{1}{\sigma L} \frac{1}{\pi^2 \Gamma\left(\delta - \frac{1}{2}\right)} \sigma \int \frac{1}{\omega} \text{p.v.} \int \frac{|\omega'|^{\delta - \frac{1}{2}} K_{\delta - \frac{1}{2}}(2|\omega'|)}{\omega - \omega'} d\omega' d\omega \leq \\ & -(\sigma - 1) \lim_{R \rightarrow \infty} \frac{y_1^{(P)}\left(\frac{\pi}{2}\right)}{y_0^{(P)}} \\ & \leq \frac{1}{\sigma L} \frac{1}{\pi^2 \Gamma\left(\delta - \frac{1}{2}\right)} (2\sigma - 1) \int \frac{1}{\omega} \text{p.v.} \int \frac{|\omega'|^{\delta - \frac{1}{2}} K_{\delta - \frac{1}{2}}(2|\omega'|)}{\omega - \omega'} d\omega' d\omega. \end{aligned}$$

These integrals can be evaluated using the formula,¹⁵

$$\int_{-\infty}^{\infty} \frac{1}{\omega} \text{p.v.} \int_{-\infty}^{\infty} \frac{f(\omega')}{\omega - \omega'} d\omega' d\omega = \pi^2 f(0),$$

which leads to

$$\frac{1}{2} \leq -(\sigma - 1) \lim_{R \rightarrow \infty} \frac{y_1^{(P)}\left(\frac{\pi}{2}\right)}{y_0^{(P)}} \leq \frac{2\sigma - 1}{2\sigma}.$$

This is, of course, consistent with the bounds (A.28) and (A.29) derived previously in a more general context.

A.12.6 Evaluating $\lim_{R \rightarrow \infty} y^{(P)}(0)$ for $\delta > \frac{1}{2}$

The logic used above to evaluate the large-space limit of $y_1^{(P)}\left(\frac{\pi}{2}\right)$ will be useful for finding the same limit of $y_1^{(P)}(0)$. The Fourier series expansion of $y_1^{(P)}(0)$ is

$$-(\sigma - 1) \frac{y_1^{(P)}(0)}{y_0^{(P)}} = \frac{1}{2} - \sum_{n \text{ even}} \frac{2\sigma - 1}{1 + (\sigma - 1) LG_{c,n}} \frac{4}{\pi^2} \sum_{m=0}^{\infty} \frac{(-1)^{\frac{n}{2}} LG_{c,2m+1}}{(2m+1)^2 - n^2}.$$

¹⁵ See, for example, equation (8.4.15) of Kanwal (1997).

Again, in the $R \rightarrow \infty$ limit, the $n = 0$ term in the sum cancels against the $1/2$.

$$-(\sigma - 1) \lim_{R \rightarrow \infty} \frac{y_1^{(P)}(0)}{y_0^{(P)}} = - \lim_{R \rightarrow \infty} \sum_{n \text{ even nonzero}} \frac{2\sigma - 1}{1 + (\sigma - 1) LG_{c,n}} \frac{4}{\pi^2} \sum_{m=0}^{\infty} \frac{(-1)^{\frac{n}{2}} LG_{c,2m+1}}{(2m+1)^2 - n^2}.$$

Splitting the sum into positive and negative contributions,

$$\begin{aligned} -(\sigma - 1) \lim_{R \rightarrow \infty} \frac{y_1^{(P)}(0)}{y_0^{(P)}} &= \lim_{R \rightarrow \infty} \sum_{\substack{n \text{ even nonzero} \\ n/2 \text{ odd}}} \frac{2\sigma - 1}{1 + (\sigma - 1) LG_{c,n}} \frac{4}{\pi^2} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2} \\ &\quad - \lim_{R \rightarrow \infty} \sum_{\substack{n \text{ even nonzero} \\ n/2 \text{ even}}} \frac{2\sigma - 1}{1 + (\sigma - 1) LG_{c,n}} \frac{4}{\pi^2} \sum_{m=0}^{\infty} \frac{LG_{c,2m+1}}{(2m+1)^2 - n^2}. \end{aligned}$$

The first sum on the right-hand side becomes

$$\frac{1}{2} \frac{1}{2\pi^2} \lim_{R \rightarrow \infty} \int \frac{2\sigma - 1}{1 + (\sigma - 1) LG_{c, \frac{\omega}{Z}}} \text{p.v.} \int \frac{LG_{c, \frac{\omega'}{Z}}}{\omega'^2 - \omega^2} d\omega' d\omega,$$

while the second sum is

$$-\frac{1}{2} \frac{1}{2\pi^2} \lim_{R \rightarrow \infty} \int \frac{2\sigma - 1}{1 + (\sigma - 1) LG_{c, \frac{\omega}{Z}}} \text{p.v.} \int \frac{LG_{c, \frac{\omega'}{Z}}}{\omega'^2 - \omega^2} d\omega' d\omega.$$

These two, of course, cancel, leading to the conclusion that

$$\lim_{R \rightarrow \infty} \frac{y_1^{(P)}(0)}{y_0^{(P)}} = 0. \tag{A.39}$$

A.12.7 Conclusion

For $\delta < \frac{1}{2}$, the inequalities (A.36) and (A.29) together imply

$$\lim_{R \rightarrow \infty} \frac{y_1^{(P)}(0)}{y_1^{(P)}\left(\frac{\pi}{2}\right)} \geq \frac{\sigma - 1}{2\sigma - 1} (1 - 2\delta).$$

This is the result presented in (1.47). For $\delta > \frac{1}{2}$, (A.39) and (A.28) give

$$\lim_{R \rightarrow \infty} \frac{y_1^{(P)}(0)}{y_1^{(P)}\left(\frac{\pi}{2}\right)} = 0.$$

This is the result (1.48).

A.13 Mathematical notation

| | |
|--|--|
| $\Gamma(x)$ | Gamma function |
| $B(x, y) \equiv \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ | Beta function |
| $B_x(p, q)$ | Incomplete beta function |
| $(a)_n \equiv \Gamma(a+n)/\Gamma(a)$ | Pochhammer symbol |
| ${}_pF_q(\alpha_1, \dots, \alpha_p; \beta_1, \dots, \beta_q; x)$ | Generalized hypergeometric function |
| $F(\alpha, \beta; \gamma; x) \equiv {}_2F_1(\alpha, \beta; \gamma; x)$ | Gauss hypergeometric function |
| ${}_2\tilde{F}_1(\alpha, \beta; \gamma; x) \equiv \frac{{}_2F_1(\alpha, \beta; \gamma; x)}{\Gamma(\gamma)}$ | Regularized hypergeometric function |
| $U(a, b, z)$ | Confluent hypergeometric function of the second kind |
| $P_\nu(x)$ | Legendre function; Legendre polynomial |
| $P_\nu^\mu(x)$ | Associated Legendre function of the first kind, LegendreP[$\nu, \mu, 3, x$] in Mathematica notation |
| $Y_l^m(\theta, \varphi)$ | Spherical harmonic function |
| $K_\nu(x)$ | Modified Bessel function of the second kind |
| $G_{p,q}^{m,n} \left(x \left \begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix} \right. \right)$ | Meijer G-function |
| $\left(\begin{matrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{matrix} \right)$ | Wigner 3j-symbol |
| p.v. \int | Cauchy principal value integral |
| δ_{nm} | Kronecker delta; 1 if $n = m$, 0 otherwise |
| f_n | Fourier coefficient of function $f(\theta)$ |
| f_l^m | Spherical harmonic coefficient of function $f(\theta, \varphi)$ |

B. APPENDIX TO CHAPTER 2

Our results in Subsection 2.3.2 assume constant returns to scale, but many of the general intuitions derived there extend to, or are even strengthened with, non-linear costs. First, consider the relationship between pass-through and quantity pass-through. Now, analogously to our analysis in that section, equilibrium is given by

$$l = - \frac{Q' \left(P - MC \left(\frac{Q - \tilde{Q}}{n} \right) - t \right)}{Q - \tilde{Q}}.$$

However, now the impact of an increase in \tilde{Q} directly on the right-hand side is

$$- \frac{\frac{Q' MC'}{n} (Q - \tilde{Q}) + Q' \left(P - MC \left(\frac{Q - \tilde{Q}}{n} \right) - t \right)}{(Q - \tilde{Q})^2} = \left(l - \frac{Q' MC'}{n} \right) \frac{1}{Q'} \frac{Q'}{Q - \tilde{Q}}.$$

Recall from above that $-\frac{Q' MC'}{n} = \frac{\epsilon_D}{\epsilon_S}$. So, by the same logic as in the main text,

$$\rho_q = \left(l + \frac{\epsilon_D}{\epsilon_S} \right) \rho. \tag{B.1}$$

Thus when there are declining returns to scale, $\rho_q > \rho$ and when they slope downward $\rho_q < \rho$. Declining returns to scale reduce pass-through and increasing returns increase it, so we can say that returns to scale have a larger impact on pass-through than on quantity pass-through, driving a wedge between them even in the monopoly case of $l = 1$.

ρ_q is always below unity if purely demand driven quantity pass-through (that which would prevail with constant returns), $\hat{\rho} \equiv \frac{l}{1-l\mu'}$ is less than 1. To see this note that

$$\rho_q - 1 = \frac{l + \frac{\epsilon_D}{\epsilon_S}}{1 + \frac{\epsilon_D}{\epsilon_S} - l\mu'} - 1 \propto l - 1 + l\mu' = l - 1 + l\mu',$$

while

$$\hat{\rho} - 1 = \frac{l - 1 + l\mu'}{1 - l\mu'} \propto l - 1 + l\mu'.$$

Thus the sign of $\rho_q - 1$ (the strategic effect, substitutes v. complements) is determined by the demand side formula. Notice that decreasing returns to scale move ρ_q towards unity (compared to the constant returns case) by increasing both the numerator and denominator while increasing returns have the opposite effect.

Finally, consider the incidence calculations. For these, it is crucial to specify the costs at which the exogenous units are produced. For this purpose, consider an alternative experiment. Rather than introducing exogenous competition, imagine the state mandating that the firms each produce their first $\frac{q}{n}$ units at marginal cost. Then equilibrium is again, as in the text given by

$$l = \frac{Q' \left[P - MC \left(\frac{Q}{n} \right) - t \right]}{Q - \tilde{Q}}$$

and thus $\frac{dQ}{dq} = l\rho$.

Profits are now $P(Q) \left(Q - \underline{q} \right) - nC \left(\frac{Q}{n} \right) + n \int_{\underline{q}=0}^{\frac{q}{n}} MC(q) dq$, so the fall in profits from an increase in \underline{q} is

$$P'l\rho \left(Q - \underline{q} \right) + P(l\rho - 1) - MC \left(\frac{Q}{n} \right) l\rho + MC \left(\frac{q}{n} \right) = -M[1 + \rho(1 - l) - \alpha].$$

where $\alpha \equiv \frac{MC(\frac{Q}{n}) - MC(\frac{q}{n})}{M}$. The argument for calculating deadweight loss incidence

proceeds exactly as before, so we obtain relative efficiency gains compared to profit losses of

$$\frac{l\rho}{1 + \rho(1 - l) - \alpha}.$$

These can be converted, just as in the text, to global incidence formulas. Notice that with decreasing returns, $\alpha < 0$ and thus deadweight loss is larger relative to profit than given by the formula in the text; when returns are increasing $\alpha > 0$ and thus deadweight loss is larger relative to profits than the formula given in the text indicates. Intuitively, the divergence is that while in the text changes in cost only affected profits, here increasing costs also affect the size of the Harberger triangle directly. Thus increasing cost has a greater impact on the $\frac{DWL}{PS}$ ratio than on the $\frac{CS}{PS}$ ratio. This is illustrated in Figure 1.

C. APPENDIX CHAPTER 3

In this appendix, we prove part (ii) of Lemma 2, namely $p'(D_{t-1}; W_t, \xi_t) \leq 0$ and $f'(D_{t-1}; W_t, \xi_t) \geq 0$. We first show that the second derivative of the value function cannot be negative. The firm's Bellman equation is

$$\begin{aligned} V(D_{t-1}; W_t, \xi_t) &= \max_{P_t} \{ (P_t - W_t) (d(P_t; \xi_t) - (1 - \delta) D_{t-1}) \\ &\quad + \beta \mathbb{E} V(d(P_t; \xi_t); W_{t+1}, \xi_{t+1}) \}. \end{aligned}$$

Denote the value of P_t that the firm actually chooses when the state variable is D_{t-1} by $P_t^{(0)}$. Now consider a small change to the initial stock of durables:

$$\begin{aligned} V(D_{t-1} + \varepsilon; W_t, \xi_t) &= \max_{P_t} \{ (P_t - W_t) (d(P_t; \xi_t) - (1 - \delta) D_{t-1} - (1 - \delta) \varepsilon) \\ &\quad + \beta \mathbb{E} V(d(P_t; \xi_t); W_{t+1}, \xi_{t+1}) \}. \end{aligned}$$

If the firm selects a wrong price, its value cannot increase:

$$\begin{aligned} V(D_{t-1} + \varepsilon; W_t, \xi_t) &\geq \left\{ (P_t^{(0)} - W_t) \left(d(P_t^{(0)}; \xi_t) - (1 - \delta) D_{t-1} - (1 - \delta) \varepsilon \right) \right. \\ &\quad \left. + \beta \mathbb{E} V(d(P_t^{(0)}; \xi_t); W_{t+1}, \xi_{t+1}) \right\}, \end{aligned}$$

$$V(D_{t-1} + \varepsilon; W_t, \xi_t) \geq - (P_t^{(0)} - W_t) (1 - \delta) \varepsilon + V(D_{t-1}; W_t, \xi_t).$$

By replacing $\varepsilon \rightarrow -\varepsilon$ we obtain another inequality of this type. Adding the two inequalities gives

$$\frac{1}{2}V(D_{t-1} + \varepsilon; W_t, \xi_t) - V(D_{t-1}; W_t, \xi_t) + \frac{1}{2}V(D_{t-1} - \varepsilon; W_t, \xi_t) \geq 0.$$

Taking the limit $\varepsilon \rightarrow 0$ we get $V''(D_{t-1}; W_t, \xi_t) \geq 0$.

The envelope theorem $V'(D_{t-1}; W_t, \xi_t) = -(1 - \delta)(p(D_{t-1}; W_t, \xi_t) - W_t)$ implies

$$V''(D_{t-1}; W_t, \xi_t) = -(1 - \delta)p'(D_{t-1}; W_t, \xi_t),$$

which means that $p'(D_{t-1}; W_t, \xi_t) \leq 0$. Given this result, and the assumption of concave utility function, (3.7) implies $d'(P_t; W_t, \xi_t) \leq 0$. Differentiating the definition of the function f gives

$$f'(D_{t-1}; W_t, \xi_t) = d'(p(D_{t-1}; W_t, \xi_t); W_t, \xi_t)p'(D_{t-1}; W_t, \xi_t); W_t, \xi_t).$$

We conclude that $f'(D_{t-1}; W_t, \xi_t) \geq 0$.

BIBLIOGRAPHY

- AGUIRRE, I., S. G. COWAN, AND J. VICKERS (2010): "Monopoly price discrimination and demand curvature," *American Economic Review*, 100(4), 1601–1615.
- AHARONY, O., S. GUBSER, J. MALDACENA, H. OOGURI, AND Y. OZ (2000): "Large N field theories, string theory and gravity," *Physics Reports*, 323(3-4), 183–386.
- ALVAREZ, F., AND R. LUCAS (2007): "General equilibrium analysis of the Eaton-Kortum model of international trade," *Journal of Monetary Economics*, 54(6), 1726–1768.
- ANDERSON, J. (1979): "A theoretical foundation for the gravity equation," *The American Economic Review*, 69(1), 106–116.
- ANDERSON, J., AND E. VAN WINCOOP (2003): "Gravity with gravitas: a solution to the border puzzle," *The American Economic Review*, 93(1), 170–192.
- (2004): "Trade costs," *Journal of Economic Literature*, 42, 691–751.
- ANDERSON, S. P., AND S. COATE (2005): "Market provision of broadcasting: a welfare analysis," *Review of Economic Studies*, 72(4), 947–972.
- ANDERSON, S. P., A. DE PALMA, AND B. KEIDER (2001): "Tax incidence in differentiated product oligopoly," *Journal of Public Economics*, 81(2), 173–192.
- ANDERSON, S. P., AND M. ENGERS (1992): "Stackelberg versus Cournot oligopoly equilibrium," *International Journal of Industrial Organization*, 10(1), 127–135.
- ANDERSON, S. P., AND R. RENAULT (2003): "Efficiency and surplus bounds in Cournot competition," *Journal of Economic Theory*, 113(2), 253–264.
- ARKOLAKIS, C., A. COSTINOT, AND A. RODRÍGUEZ-CLARE (forthcoming): "New trade models, same old gains?," *The American Economic Review*.
- ARMINGTON, P. (1969): "A theory of demand for products distinguished by place of production," *IMF Staff papers*, 17, 159–176.
- ARMSTRONG, M. (2006): "Competition in two-sided markets," *RAND Journal of Economics*, 37(3), 668–691.

- ARMSTRONG, M., AND J. VICKERS (2010): "A model of delegated project choice," *Econometrica*, 78(1), 213–244.
- ARMSTRONG, M., J. VICKERS, AND J. ZHOU (2009): "Prominence and consumer search," *RAND Journal of Economics*, 40(2), 209–233.
- ARMSTRONG, M., AND J. ZHOU (2011): "Paying for prominence," *Economic Journal*, 121(556), F368–F395.
- ATKIN, D., AND D. DONALDSON (2012): "Who's getting globalized? Intra-national trade costs and world price pass-through in South Asia and Sub-Saharan Africa," This paper is under preparation.
- BAK, P., K. CHEN, J. SCHEINKMAN, AND M. WOODFORD (1993): "Aggregate fluctuations from independent sectoral shocks: self-organized criticality in a model of production and inventory dynamics," *Ricerche Economiche*, 47(1), 3–30.
- BAKER, J. B., AND T. F. BRESNAHAN (1985): "The gains from merger or collusion in product-differentiated industries," *Journal of Industrial Economics*, 33(4), 427–444.
- BALDENIUS, T., AND S. REICHELSTEIN (2000): "Comparative statics of monopoly pricing," *Economic Theory*, 16(2), 465–469.
- BALDWIN, R., R. FORSLID, P. MARTIN, G. OTTAVIANO, AND F. ROBERT-NICOUD (2005): *Economic geography and public policy*. Princeton University Press.
- BALISTRERI, E., R. HILLBERRY, AND T. RUTHERFORD (2011): "Structural estimation and solution of international trade models with heterogeneous firms," *Journal of International Economics*, 83(2), 95–108.
- BARSKY, R. B., C. L. HOUSE, AND M. S. KIMBALL (2007): "Sticky-price models and durable goods," *American Economic Review*, 97(3), 984–998.
- BECKER, G. S. (1991): "A note on restaurant pricing and other examples of social influences on price," *Journal of Political Economy*, 99(5), 1109–1116.
- BECKER, G. S., AND K. M. MURPHY (1993): "A simple theory of advertising as a good or a bad," *Quarterly Journal of Economics*, 108(4), 941–964.
- BEHRENS, K., C. ERTUR, AND W. KOCH (2007): "Dual gravity: using spatial econometrics to control for multilateral resistance," *Journal of Applied Econometrics*.
- BERNARD, A., J. JENSEN, S. REDDING, AND P. SCHOTT (2007): "Firms in international trade," *The Journal of Economic Perspectives*, 21(3), 105–130.

- BOND, E. W., AND L. SAMUELSON (1984): "Durable good monopolies with rational expectations and replacement sales," *RAND Journal of Economics*, 15(3), 336–345.
- BOWLEY, A. L. (1924): *The mathematical groundwork of economics*. Oxford University Press.
- BRANDER, J. A., AND B. J. SPENCER (1981): "Tariffs and the extraction of foreign monopoly rents under potential entry," *The Canadian Journal of Economics*, 14(3), 371–389.
- BREITENLOHNER, P., AND D. FREEDMAN (1982a): "Positive energy in anti-de Sitter backgrounds and gauged extended supergravity," *Physics Letters B*, 115(3), 197–201.
- (1982b): "Stability in gauged extended supergravity," *Annals of Physics*, 144(2), 249–281.
- BULOW, J., AND P. KLEMPERER (Forthcoming): "Regulated prices, rent-seeking and consumer surplus," *Journal of Political Economy*.
- BULOW, J. I. (1982): "Durable-goods monopolists," *Journal of Political Economy*, 90(2), 314–332.
- BULOW, J. I., J. D. GEANAKOPOLOS, AND P. D. KLEMPERER (1985): "Multimarket oligopoly: strategic substitutes and compliments," *Journal of Political Economy*, 93(3), 488–511.
- BULOW, J. I., AND P. PFLEIDERER (1983): "A note on the effect of cost changes on prices," *Journal of Political Economy*, 91(1), 182–185.
- CAPLIN, A., AND J. LEAHY (2006): "Equilibrium in a durable goods market with lumpy adjustment," *Journal of Economic Theory*, 128(1), 187–213.
- CHAMBERLIN, E. H. (1931): *The theory of monopolistic competition*. Cambridge, MA: Harvard University Press.
- CHANNEY, T. (2008): "Distorted gravity: the intensive and extensive margins of international trade," *The American Economic Review*, 98(4), 1707–1721.
- (2011a): "The gravity equation in international trade: an explanation," <https://sites.google.com/site/thomaschaney/research>.
- (2011b): "The network structure of international trade," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1758451.
- CHETTY, R. (2009): "Sufficient statistics for welfare analysis: a bridge between structural and reduced-form methods," *Annual Review of Economics*, 1, 451–488.

- COASE, R. H. (1972): "Durability and monopoly," *Journal of Law and Economics*, 15(1), 143–49.
- COURNOT, A. A. (1838): *Recherches sur les principes mathematiques de la theorie des richesses*. Paris.
- COWAN, S. (2010): "Third-degree price discrimination and consumer surplus," <http://ideas.repec.org/p/oxf/wpaper/462.html>.
- DIXIT, A., AND V. NORMAN (1978): "Advertising and welfare," *Bell Journal of Economics*, 9(1), 1–17.
- EATON, J., AND S. KORTUM (2002): "Technology, geography, and trade," *Econometrica*, 70(5), 1741–1779.
- (in progress): *Technology and the global economy: a framework for quantitative analysis*. Princeton University Press, manuscript available online.
- EDELMAN, B., M. OSTROVSKY, AND M. SCHWARZ (2007): "Internet advertising and the generalized second price auction: selling billions of dollars worth of keywords," *American Economic Review*, 97(1), 242–259.
- ENGEL, C., AND J. ROGERS (1996): "How wide is the border?," *The American Economic Review*, 86(5), 1112–1125.
- ESTEBAN, S., AND M. SHUM (2007): "Durable-goods oligopoly with secondary markets: the case of automobiles," *RAND Journal of Economics*, 38(2).
- FABINGER, M., AND E. G. WEYL (2012): "Incidence, demand forms and imperfect competition," This work is in progress.
- FARRELL, J. (2008): "Some welfare analytics of aftermarket," http://www.law.northwestern.edu/searlecenter/papers/Farrell_Analytics.pdf.
- FARRELL, J., AND C. SHAPIRO (2010): "Antitrust evaluation of horizontal mergers: an economic alternative to market definition," *Berkeley Electronic Journal of Theoretical Economics: Policies and Perspectives*, 10(1).
- FUJITA, M., P. KRUGMAN, AND A. VENABLES (1999): *The spatial economy: cities, regions and international trade*. MIT Press.
- GABAIX, X., D. LAIBSON, D. LI, H. LI, S. RESNICK, AND C. G. DE VRIES (2010): "Extreme value theory and the equilibrium prices in large economies," <http://pages.stern.nyu.edu/~xgabaix/papers/CompetitionEVT.pdf>.

- GABAIX, X., AND D. I. LAIBSON (2006): "Shrouded attributes, consumer myopia, and information suppression in competitive markets," *Quarterly Journal of Economics*, 121(2), 505–540.
- GILBERT, R., AND C. SHAPIRO (1990): "Optimal patent length and breadth," *RAND Journal of Economics*, 21(1), 106–112.
- GOMES, R. (2011): "Optimal auction design in two-sided markets," <http://sites.google.com/site/northwesternrenatogomes/Home/research>.
- GOMES, R., AND A. PAVAN (2011): "Price discrimination in many-to-many matching markets," http://faculty.wcas.northwestern.edu/~apa522/matching_design.pdf.
- GOOS, M., P. V. CAYSEELE, AND B. WILLEKENS (2011): "Platform pricing in matching markets," <https://lirias.kuleuven.be/bitstream/123456789/328974/1/DPS1132.pdf>.
- GOPINATH, G., P. GOURINCHAS, C. HSIEH, AND N. LI (forthcoming): "International prices, costs and markup differences," *The American Economic Review*.
- GORODNICHENKO, Y., AND L. TESAR (2009): "Border effect or country effect? Seattle may not be so far from Vancouver after all," *American Economic Journal: Macroeconomics*, 1(1), 219–241.
- GRADSHTEYN, I., AND I. RYZHIK (2007): *Table of integrals, series, and products*. Academic Press, seventh edn.
- GUL, F., H. SONNENSCHN, AND R. WILSON (1986): "Foundations of dynamic monopoly and the Coase conjecture," *Journal of Economic Theory*, 39(1), 155–190.
- HARBERGER, A. C. (1954): "Monopoly and resource allocation," *American Economic Review Papers and Proceedings*, 44(2), 77–87.
- HELPMAN, E., AND P. KRUGMAN (1987): *Market structure and foreign trade: increasing returns, imperfect competition, and the international economy*. The MIT Press.
- HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): "Estimating trade flows: trading partners and trading volumes," *The Quarterly Journal of Economics*, 123(2), 441–487.
- HELPMAN, E., M. MELITZ, AND S. YEAPLE (2004): "Export versus FDI with heterogeneous firms," *The American Economic Review*, 94(1), 300–316.
- HILLBERRY, R., AND D. HUMMELS (2008): "Trade responses to geographic frictions: a decomposition using micro-data," *European Economic Review*, 52(3), 527–550.

- HOLMSTRÖM, B., AND P. MILGROM (1991): "Multitask principal-agent analyses: incentive contracts, asset ownership, and job design," *Journal of Law, Economics, and Organization*, 7(Special Issue), 24–52.
- HUMMELS, D. (2001): "Toward a geography of trade costs," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=160533.
- JAFFE, S., AND E. G. WEYL (2012): "The first-order approach to merger analysis," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1765024.
- JEHIEL, P., AND B. MOLDOVANU (2001): "Efficient design with interdependent valuations," *Econometrica*, 69(5), 1237–1259.
- JOHNSON, H. G. (1953–1954): "Optimum tariffs and retaliation," *Review of Economic Studies*, 21(2), 142–153.
- JOHNSON, J. P., AND D. P. MYATT (2006): "On the simple economics of advertising, marketing, and product design," *American Economic Review*, 96(3), 756–784.
- KANWAL, R. (1997): *Linear integral equations*. Birkhäuser, second edn.
- KLEMPERER, P. D., AND M. A. MEYER (1989): "Supply function equilibria in oligopoly under uncertainty," *Econometrica*, 57(6), 1243–1277.
- KRUGMAN, P. (1980): "Scale economies, product differentiation, and the pattern of trade," *The American Economic Review*, 70(5), 950–959.
- LAFFONT, J.-J., AND J. TIROLE (1993): *A theory of incentives in procurement and regulation*. Cambridge, MA: MIT Press.
- MANKIW, N. G., AND M. D. WHINSTON (1986): "Free entry and social inefficiency," *RAND Journal of Economics*, 17(1), 48–58.
- MARES, V., AND J. M. SWINKELS (2011): "On the analysis of asymmetric first price auctions," <http://www.kellogg.northwestern.edu/faculty/swinkels/Homepage/Papers/Main%20Paper%205-4-11.pdf>.
- MARSHALL, A. (1890): *Principles of economics*. New York: Macmillan.
- MARTIN, P., AND C. ROGERS (1995): "Industrial location and public infrastructure," *Journal of International Economics*, 39(3-4), 335–351.
- MATSUYAMA, K. (1999): "Geography of the world economy," <http://econpapers.repec.org/paper/nwucmsems/1239.htm>.
- MCCALLUM, J. (1995): "National borders matter: Canada-US regional trade patterns," *The American Economic Review*, 85(3), 615–623.

- MELITZ, M. (2003): "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, 71(6), 1695–1725.
- MELITZ, M., AND G. OTTAVIANO (2008): "Market size, trade, and productivity," *Review of Economic studies*, 75(1), 295–316.
- MYATT, D. P., AND E. B. RASMUSEN (2011): "Posted prices vs. haggling: the economics of isoperfect price discrimination," <http://rasmusen.org/papers/pdisc-myatt-rasmusen.pdf>.
- NASH, J. F. J. (1950): "The bargaining problem," *Econometrica*, 18(2), 155–162.
- NOCKE, V., AND M. D. WHINSTON (2011): "Merger policy with merger choice," <http://www.wcas.northwestern.edu/csio/Papers/2010/CSIO-WP-0112.pdf>.
- OKAWA, Y., AND E. VAN WINCOOP (2010): "Gravity in international finance," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1628043.
- OSSA, R. (2011): "A "New Trade" Theory of GATT/WTO negotiations," *Journal of Political Economy*, 119(1), 122–152.
- OTTAVIANO, G., T. TABUCHI, AND J. THISSE (2002): "Agglomeration and trade revisited," *International Economic Review*, 43(2), 409–435.
- PIGOU, A. C. (1920): *The economics of welfare*. London: MacMillan.
- QUINT, D. (2012): "Imperfect competition with complements and substitutes," <http://www.ssc.wisc.edu/~dquint/papers/quint-complements-substitutes.pdf>.
- RENY, P., S. WILKIE, AND M. A. WILLIAMS (2011): "Tax incidence under imperfect competition," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1874346.
- ROBINSON, J. (1933): *The economics of imperfect competition*. London: MacMillan.
- ROCHET, J.-C., AND L. A. STOLE (2003): "The economics of multidimensional screening," in *Advances in Economics and Econometrics: Theory and Applications*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, vol. 1 of *Eighth World Congress*. Cambridge University Press.
- ROCHET, J.-C., AND J. TIROLE (2003): "Platform competition in two-sided markets," *Journal of the European Economic Association*, 1(4), 990–1029.
- (2006): "Two-sided markets: a progress report," *RAND Journal of Economics*, 37(3), 645–667.
- ROMALIS, J. (2004): "Factor proportions and the structure of commodity trade," *American Economic Review*, pp. 67–97.

- ROSSI-HANSBERG, E. (2005): "A spatial theory of trade," *American Economic Review*, 95(5), 1464–1491.
- RYSMAN, M. (2009): "The economics of two-sided markets," *Journal of Economic Perspectives*, 23(3), 125–43.
- SAEZ, E. (2001): "Using elasticities to derive optimal income tax rates," *The Review of Economic Studies*, 68(1), 205–229.
- SCHMALENSEE, R. (1981): "Output and welfare implications of monopolistic third-degree price discrimination," *American Economic Review*, 71(1), 242–247.
- SHAPIRO, C. (1996): "Mergers with differentiated products," *Antitrust*, 10, 23–30.
- SONNENSCHN, H. (1968): "The dual of duopoly is complimentary monopoly: or, two of Cournot's theories are one," *Journal of Political Economy*, 76(2), 316–318.
- SPENCE, A. M. (1975): "Monopoly, quality, and regulation," *Bell Journal of Economics*, 6(2), 417–429.
- SPENGLER, J. J. (1950): "Vertical integration and antitrust policy," *Journal of Political Economy*, 50(4), 347–352.
- STOKEY, N., R. LUCAS, AND E. PRESCOTT (1989): *Recursive methods in economic dynamics*. Harvard University Press.
- STOKEY, N. L. (1981): "Rational expectations and durable goods pricing," *Bell Journal of Economics*, 12(1), 112–128.
- TIROLE, J. (1988): *The theory of industrial organization*. MIT Press.
- TURNER, L. J. (2012): "The incidence of student financial aid: evidence from the Pell Grant program," <http://www.columbia.edu/~ljt2110/LTurner-JMP.pdf>.
- VARIAN, H. R. (1985): "Price discrimination and social welfare," *American Economic Review*, 75(4), 870–875.
- VEIGA, A., AND E. G. WEYL (2012): "Multidimensional product design," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1935912.
- VILLAS-BOAS, S. B., AND R. HELLERSTEIN (2006): "Identification of supply models of retailer and manufacturer oligopoly pricing," *Economics Letters*, 90(1), 132–140.
- VON STACKELBERG, H. F. (1934): *Marktform und Gleichgewicht*. Vienna: Julius Springer.
- WALDMAN, M. (2003): "Durable goods theory for real world markets," *Journal of Economic Perspectives*, 17(1), 131–154.

- WERDEN, G. J. (1996): "A robust test for consumer welfare enhancing mergers among sellers of differentiated products," *Journal of Industrial Economics*, 44(4), 409–413.
- WEYL, E. G. (2009): "Monopoly, Ramsey and Lindahl in the Rochet and Tirole (2003)," *Economics Letters*, 103(2), 99–100.
- (2010): "A price theory of multi-sided platforms," *American Economic Review*, 100(4), 1642–1672.
- WEYL, E. G., AND M. FABINGER (2009): "Pass-through as an economic tool," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1324426.
- WEYL, E. G., AND J. TIROLE (2012): "Market power screens willingness-to-pay," http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1659449.